М. БОРН, Э. ВОЛЬФ

ОСНОВЫ ОПТИКИ



PRINCIPLES OF OPTICS

Electromagnetic Theory of Propagation, Interference and Diffraction of Light

by

MAX BORN

M. A., Dr. Phil., F. R. S.

Nobel Laureate

Professor Emeritus at the Universities of Göttingen and Edinburgh

and

EMIL WOLF

Ph. D., D. Sc.

Professor of Physics, University of Rochester, N. Y.

with contributions by

A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes,
A. M. Taylor, P. A. Wayman and W. L. Wilcock

FOURTN EDITION

PERGAMON PRESS

OXFORD·LONDON·EDINBURGH·NEW YORK
PARIS·FRANKFURT

1968

М. БОРН, Э. ВОЛЬФ

ОСНОВЫ ОПТИКИ

ИЗДАНИЕ 2-е, ИСПРАВЛЕННОЕ

Перевод с английского
С. Н. БРЕУСА, А. И. ГОЛОВАШКИНА, А. А. ШУБИНА

Под редакцией
Г. П. МОТУЛЕВИЧ

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1973

535
Б Д82
УДК 535

Основы оптики. Борн М, Вольф Э., изд. 2-е. Перевод с английского. Главная редакция физико-математической литературы изд-ва «Наука», 1973.

Наиболее полный и авторитетный труд по оптике в мировой литературе, учитывающий все последние достижения классической теории.

Излагаются макроскопические уравнения Максвелла с формально введенными константами и подробно разбираются вопросы распространения электромагнитных волн в среде, а также связь этих констант с поляризацией и намагничением.

Уравнения геометрической оптики последовательно выводятся из уравнении Максвелла (при этом автоматически учитывается поперечность и векторный характер световых волн) и затем применяются к теории оптического изображения и к расчету аберраций. Рассматриваются интерференция, элементарная и строгая теория дифракции, дифракционная теория аберраций и дифракция света на ультразвуковых волнах. Подробно излагаются вопросы распространения, интерференции и дифракции частично когерентного света; основное внимание уделяется случаю квазимонохроматического излучения, причем общее рассмотрение строится на использовании метода корреляционных функций Излагаются вопросы металлооптики и кристаллооптики.

Во всей книге много внимания уделяется изложению математического аппарата.

Таблиц 30, иллюстраций 369, библиография 814 назв.

Б $\frac{0234-1810}{042\ (02)\text{-}73}$ 108-73

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Переводимая книга крупных ученых М. Борна и Э. Вольфа задумана ими как первый том фундаментального курса по теоретической оптике, в котором разбираются вопросы распространения электромагнитных волн, интерференция и дифракция света, геометрическая оптика, металлооптика и кристаллооптика. Молекулярную и квантовую оптику авторы предполагали изложить в последующих томах.

Данный том — одна из последних книг, написанная с участием М. Борна, лауреата Нобелевской премии по физике 1954 г. Недавняя кончина (5 января 1970 г.) этого крупного физика, автора многочисленных научных работ и книг, создателя большой школы теоретической физики, не позволила ему лично довести задуманный план до конца.

Советским физикам работы М. Борна хорошо известны. Его книга «Оптика», вышедшая в русском переводе в 1935 г., до последнего времени была одной из настольных книг физиков-оптиков. Настоящая книга существенно отличается от прежней. Материал изложен более последовательно и полно, добавлены новые разделы по актуальным вопросам оптики.

В главе 1 рассматриваются макроскопические уравнения Максвелла с формально введенными диэлектрической и магнитной проницаемостями и электропроводностью, и подробно разбираются вопросы распространения электромагнитных волн. Связь введенных констант с поляризацией и намагничением вводится в главе 2, где затронуты также некоторые вопросы молекулярной оптики и дается элементарная теория дисперсии света. В дальнейшем на систематическом применении уравнений Максвелла, рассмотренных в главах 1 и 2, строится вся книга.

Главы 3—6 посвящены геометрической оптике, изложение которой оригинально и интересно. Уравнения геометрической оптики последовательно выводятся из уравнений Максвелла. При этом автоматически учитывается поперечность и векторный характер световых волн. Далее полученные уравнения применяются к теории оптического изображения и к расчету аберраций оптических систем. Рассмотрению указанных вопросов в книге не случайно отведено много места, что отражает успехи, достигнутые за последнее время в геометрической оптике.

Главы 7—12 посвящены интерференции и дифракции света. В главе 7 рассматриваются явление интерференции и его применение в интерференционных приборах, а в главе 8 дается элементарная теория дифракции. Строгая теория дифракции, основанная на уравнениях Максвелла и соответствующих граничных условиях, приводится в главе 11. Эта теория используется для решения задач дифракции света на идеально проводящих плоском экране и полуплоскости, а также для некоторых других задач. В главе 9 дается дифракционная теория аберраций. Разбираются искажения дифракционного изображения точечных и протяженных источников, вызванные аберрациями. В главе 12 рассматривается дифракция света на ультразвуковых волнах, которая обычно почти не освещается. Очень интересна глава 10, посвященная распространению, интерференции и дифракции частично коге-

рентного света. Особенно подробно обсуждается случай квазимонохроматического излучения. Общее рассмотрение строится на использовании техники корреляционных функций. Рассмотрены некоторые весьма интересные приложения: вопрос о когерентности, возникающей из-за наличия оптической системы при некогерентном источнике, о влиянии освещения на разрешающую систему микроскопа, о формировании изображений при наличии частично когерентного квазимонохроматического света, о поляризационных свойствах квазимонохроматического света. Появление такой главы в курсе классической оптики представляется очень целесообразным.

Главы 13 и 14, посвященные металлооптике и кристаллооптике, мало отличаются от соответствующих глав старого курса и не отражают тех изменений, которые произошли в этих областях за время, прошедшее после издания прежнего курса.

В книге много места уделяется изложению математического аппарата, необходимого при рассмотрении затронутых вопросов. Помимо математических приложений в конце книги, математические отступления делаются во многих главах, что несколько удлиняет текст, но делает изложение более понятным.

Мы не сомневаемся в том, что это фундаментальное руководство по оптике М. Борна и Э. Вольфа будет (так же как и книга М. Борна «Оптика») долгие годы настольной книгой всех физиков, занимающихся оптикой.

Книга переведена полностью, без сокращений. Поскольку в списке литературы, приведенном авторами, не отражены работы советских ученых, мы снабдили книгу примечаниями, включающими соответствующие ссылки.

Книгу переводили: С. Н. Бреус (введение, главы 3—5, 9, приложения), А. И. Головашкин (главы 1, 2, 10, 13, 14) и А. А. Шубин (главы 6—8, 11, 12).

Г. Мотулевич

ПРЕДИСЛОВИЕ К ЧЕТВЕРТОМУ ИЗДАНИЮ

Ввиду внушительного объема этой книги было сочтено нецелесообразным включать в это новое издание дополнительный материал, относящийся к последним достижениям в оптике. Тем не менее, мы сделали ряд новых исправлений и улучшений в тексте и добавили ссылки на некоторые недавние публикации.

Мы обязаны д-ру Марчанду и Кусакаве за предоставление списков опечаток и ошибок, замеченных в предыдущих изданиях. Мы также признательны д-ру Бедарду за помощь при переработке.

Бэд Пирмонт и Рочестер
Август 1968

Макс Борн
Эмиль Вольф

ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ

Целый ряд ошибок и опечаток, имевшихся в более ранних изданиях этой книги, были исправлены, а также были добавлены ссылки на некоторые недавние публикации.

Был также включен новый рисунок (8. 54), относящийся к последним достижениям в технике восстановления волновых фронтов. Мы признательны Лейту и Юпетниксу за предоставление оригинальных фотографий.

Бэд Пирмонт и Рочестер
Июль 1965

Макс Борн
Эмиль Вольф

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

В процессе подготовки нового издания этой книги был исправлен ряд ошибок и опечаток, сделано несколько небольших добавлений и добавлены новые ссылки.

После выхода в свет первого издания и почти точно три года назад появились первые оптические мазеры (лазеры). С их помощью можно получать очень интенсивные высококогерентные пучки света. Хотя очевидно, что оптические мазеры будут иметь огромное значение не только в оптике, но и в других областях науки и техники, в новом издании не сделано никакого обзора, касающегося их действия Дело в том, что принцип их действия выходит за рамки классической электромагнитной теории, на основе которой излагается большая часть материала книги. Мы, однако, включили несколько ссылок как на работы, в которых использовался свет, генерируемый лазерами, так и на те, которые появились благодаря потенциальным возможностям этих новых оптических приборов.

Мы признательны многочисленным читателям, которые указали нам на некоторые ошибки и опечатки. Мы также благодарны д-ру Карчевски и Мета за помощь при переработке.

Бэд Пирмонт и Рочестер
Ноябрь 1962

Макс Борн
Эмиль Вольф

ИЗ ПРЕДИСЛОВИЯ К ПЕРВОМУ ИЗДАНИЮ

Мысль о написании этой книги родилась в связи с многочисленными запросами относительно возможности опубликования на английском языке книги по оптике, написанной одним из нас*) более 25 лет назад. Предварительное ознакомление с литературой показало, что за истекшее время появилось множество исследований почти во всех областях оптики, и старая книга перестала давать о ней исчерпывающее и гармоничное представление. Таким образом, мы увидели, что перевод нецелесообразен, и была написана совершенно новая книга, которую мы предлагаем теперь читателю. В процессе ее подготовки вскоре стало очевидным, что если даже включить в нее только наиболее важные результаты, полученные с момента опубликования «Оптики», книга окажется непомерно большой. Поэтому мы были вынуждены посвятить ее более узкой тематике. Старая книга также не охватывала всю оптику. В ней не рассматривалась оптика движущихся сред, оптика рентгеновских и γ-лучей, теория спектров и связь между оптикой и атомной физикой; в ней не обсуждалось также действие света на наш орган зрения — глаз. Изучение этих вопросов более уместно в таких разделах науки, как теория относительности, квантовая механика, атомная и ядерная физика, физиология. В настоящей книге не обсуждаются не только эти вопросы, но и исключена классическая молекулярная оптика, которой была посвящена почти половина немецкого издания. Таким образом, наше рассмотрение ограничивается оптическими явлениями, допускающими исследование в рамках феноменологической теории Максвелла. Сюда относятся все явления, в которых атомистическое строение вещества не играет существенной роли. О связи с атомной физикой, квантовой механикой и физиологией говорится лишь по ходу изложения в кратких отступлениях. Несмотря на такие ограничения, книга оказалась значительно больше «Оптики», что дает представление о размахе исследований по классической оптике за последнее время.

Мы стремились к тому, чтобы в рамках указанных ограничений дать достаточно полное представление о современном состоянии этой области физики и старались изложить теорию таким образом, чтобы можно было проследить вывод практически всех результатов из основных уравнений электромагнитной теории Максвелла, с описания которой начинается наша книга.

В гл. 1 обсуждаются основные свойства электромагнитного поля и формально (с помощью обычных материальных постоянных) описывается влияние вещества на распространение электромагнитных возмущений. В гл. 2 развивается более физический подход к изучению его влияния; показано, что при наличии внешнего поля каждый элементарный объем среды можно считать источником вторичной (рассеянной) элементарной волны и комбинация этих волн образует наблюдаемое макроскопическое поле. Такой подход имеет важное физическое значение, и его мощь иллюстрируется в одной из последующих глав (гл. 12) при исследовании дифракции света на ультразвуковых волнах,

*) M. Born, Optik, Springer, Berlin, 1933. (М. Борн, Оптика, ГНТИУ, Харьков — Киев, 1937.)

впервые рассмотренной таким образом Бхатиа и Ноблом (гл. 12 написана проф. Бхатиа).

Значительная часть гл. 3 посвящена доказательству того, что геометрическая оптика является следствием волновой теории Максвелла в предельном случае коротких длин волн. Кроме обсуждения основных свойств лучей и волновых фронтов рассматриваются также вопросы, связанные с их векторными свойствами (изменение направлений векторов полей при распространении). Подробное обсуждение основ геометрической оптики казалось нам желательным в связи с большими успехами, достигнутыми за последнее время в микроволновой оптике (оптике коротких радиоволн). Они нередко становились возможными благодаря сходству геометрической и микроволновой оптик и обеспечили создание новой экспериментальной техники для проверки предсказаний теории. Мы сочли целесообразным вынести математический аппарат геометрической оптики (вычисление вариаций) из основного текста; его изложение (см. приложение 1) основано главным образом на неопубликованных лекциях Гильберта, прочитанных им в Геттингенском университете в начале века. В следующем приложении (приложение 2), написанном проф. Габором, раскрывается близкая формальная аналогия между геометрической оптикой, классической и волновой механиками и электронной оптикой, проявляющаяся при обсуждении этих разделов на языке вариационного исчисления.

Мы не жалеем, что наше рассмотрение геометрической теории изображений (гл. 4) основывается на классическом методе характеристических функций Гамильтона. Этот метод не нашел широкого применения при конструировании оптических инструментов, но он, тем не менее, очень удобен для представления различных аспектов теории с единой точки зрения. Несомненно, некоторые результаты можно получить проще с помощью специальных предположений; однако, хотя последний способ удобен при решении некоторых частных задач, он может иметь лишь иллюстративное значение в книге, посвященной систематическому построению теории из нескольких простых постулатов.

Дефекты оптических изображений (влияние аберраций) можно исследовать либо в рамках геометрической оптики (когда аберрации велики), либо в рамках теории дифракции (когда аберрации достаточно малы). Раньше обычно возникали трудности при попытках сравнить результаты этих двух подходов, поскольку исходные положения последних совершенно различны. Мы попытались развить более единообразный метод, основанный на понятии о деформации волновых фронтов. При изложении геометрической теории аберраций (гл. 5) мы нашли возможным и целесообразным использовать старый метод Шварцшильда после небольшого изменения введенного им эйконала. В главе, посвященной дифракционной теории аберраций (гл. 9), дается обзор теории Нижбера — Цернике; в ней излагается также вводный раздел об изображении при когерентном и некогерентном освещении протяженных объектов, где используются в основном преобразования Фурье.

В гл. 6, написанной Вейманом, приводится краткое описание основных оптических систем, формирующих изображение. Это делается с целью подготовить читателя к изучению тех разделов книги, в которых рассматривается теория формирования изображений.

Гл. 7 посвящена теории интерференции и интерферометрам. В соответствующих разделах старой «Оптики» излагались некоторые основные положения теории, затронутые в настоящем издании, однако эти разделы полностью переписаны и значительно расширены Уилкоком.

В гл. 8 излагается в основном дифракционная теория Френеля — Кирхгофа и некоторые ее приложения. Кроме обычных вопросов, в ней содержится подробное обсуждение главной проблемы теории формирования оптических изображений — анализа пространственного распределения света вблизи гео-

метрического фокуса. Дается также обзор менее известного подхода к решению задач дифракции, основанного на понятии поверхностной дифрагировавшей волны Юнга.

В перечисленных выше главах свет считался почти всегда монохроматическим (а следовательно, полностью когерентным), исходящим из точечного источника. В гл. 10 рассматривается более реальный случай, когда свет испускается источником конечных размеров и его длины волн (частоты) лежат в конечном интервале. Этот случай обсуждается в теории частичной когерентности, получившей очень существенное развитие в последние годы. Фактически строгая теория интерференции и дифракции частично когерентного света создана лишь в настоящее время. Кроме того, в гл. 10 рассматривается вопрос о частичной поляризации, трактуемый с точки зрения теории когерентности.

Гл. 11 посвящена строгой теории дифракции, получившей за последние 20 лет огромное развитие *), вызванное в основном прогрессом в ультракоротковолновой радиотехнике. Эта глава, а также приложение 3, посвященное математическим методам наибыстрейшего спуска и стационарной фазы, написаны Клеммовым.

Изложение последних двух глав, «Металлооптика» (гл. 13) и «Кристаллооптика» (гл. 14), основано на соответствующих главах старого издания «Оптики», однако они существенно пересмотрены и дополнены соответственно Тейлором и Стоксом. Вероятно, этот материал обсуждается не так подробно, как он того заслуживает. Однако строгое рассмотрение металлооптики возможно только с привлечением квантовой механики электронов, что выходит за рамки настоящей книги. В кристаллооптике «центр тяжести» постепенно сместился от видимого излучения к рентгеновскому, и прогресс, достигнутый в этой области за последние годы, имеет не теоретическое, а техническое значение.

Хотя мы пытались создать книгу, сходную по своим методам и общему подходу со старой «Оптикой», очевидно, что настоящую книгу нельзя считать ни переводом «Оптики», ни полной компиляцией известных сведений. Что касается нашего личного участия в создании этой книги, то следует сказать, что старший соавтор (Макс Борн) включил в нее те разделы «Оптики», которые послужили основой для ряда глав настоящего издания, и принимал активное участие в создании общего плана книги и в многочисленных дискуссиях, посвященных спорным вопросам, характеру изложения и т. д. Большая часть работы по сбору материала, написанию и исправлению текста была проделана младшим соавтором (Эмилем Вольфом).

Естественно, мы стремились пользоваться единообразными обозначениями во всем тексте. Однако для книги, посвященной такой обширной области, число букв в известных алфавитах слишком мало. Поэтому нам не всегда удавалось применять наиболее изящные обозначения, но мы надеемся, что по крайней мере в каждой главе разные величины не обозначаются одинаковыми буквами.

Как правило, мы использовали векторные обозначения, общепринятые в Великобритании. После многочисленных сомнений мы все же отказались от использования одного только оператора набла (∇) и ввели обычные div, grad и rot. Мы не приняли также современных электротехнических единиц, так как их преимущество проявляется лишь при описании чисто электромагнитных измерений, занимающих ничтожное место в нашей книге; более того, мы надеемся, что если когда-нибудь будет написан второй том («Молекулярная и атомная оптика»), а возможно, и третий («Квантовая оптика»), то к тому вре-

*) В обзорной статье Баукампа (C. J. Bouwkamp, в сб. «Rep. Progr. Phys.», Phys. Soc. London, 1954, Vol. 17, p. 35) имеется более 500 ссылок на работы, выполненные за период 1940—1954 гг.

мени система СГС, используемая в теоретической физике, снова окажется в почете. Хотя в этой системе единиц магнитная проницаемость μ для большинства сред практически не отличается от единицы в оптическом диапазоне частот, мы все же оставляли ее в некоторых уравнениях. В результате уравнения оказываются более симметричными и позволяют получать «двойные результаты», пользуясь свойствами симметрии уравнений Максвелла. Для периодических во времени полей мы использовали во всей книге (при комплексном представлении) множитель exp $(-i\omega t)$.

Мы не собирались ссылаться в тексте на все подходящие публикации. Приведенные ссылки, которые, как мы надеемся, относятся к наиболее важным работам, служат для того, чтобы помочь читателю ориентироваться в литературе; отсутствие ссылки на какую-нибудь работу не должно истолковываться как недооценка ее авторами.

В заключение мы хотели поблагодарить многих друзей и коллег за советы и помощь. В первую очередь мы выражаем благодарность проф. Габору за полезные советы и помощь на ранней стадии работы, а также за предоставление черновика рукописи, в которой излагается его остроумный метод восстановленных волновых фронтов (см. § 8.10). Мы также глубоко благодарны д-ру Абеле, подготовившему черновик текста, послужившего основой при написании § 1.6 и посвященного вопросу о распространении электромагнитных волн через слоистые среды, вопросу, в решение которого сам д-р Абеле внес существенный вклад. Кроме того, мы обязаны д-ру Биллингсу за советы, касающиеся этой же проблемы

Мы очень благодарны д-ру Гопкинсу, д-ру Сильверману, д-ру Велфорду и д-ру Вилли за критические замечания и ценные советы; мы признательны им, а также д-ру Блеку, д-ру Браддику, д-ру Чако, д-ру Кану, Нисбету, д-ру Россу и Силлитто за внимательное чтение различных разделов рукописи.

Бэд Пирмонт и Манчестер
Январь 1959

Макс Борн
Эмиль Вольф

ИСТОРИЧЕСКОЕ ВВЕДЕНИЕ

Физические принципы, лежащие в основе оптических явлений, обсуждению которых посвящена настоящая книга, были в основном сформулированы до 1900 г.

В начале XX века оптика, как и вся остальная физика, подверглась глубокому революционному переосмысливанию, вызванному открытием кванта энергии. Однако это открытие, коренным образом изменившее наши представления о природе света, не сделало старые теории ненужными, но позволило выявить их ограниченность и установить пределы их применимости. Поэтому обобщение старых принципов и методов и их использование в различных конкретных случаях продолжались и продолжаются столь же интенсивно и в настоящее время.

При попытке систематического изложения знаний, приобретенных в течение нескольких столетий в такой обширной области, как оптика, невозможно придерживаться истории, изобиловавшей ошибочными идеями и отступлениями. Поэтому в этой вводной главе мы считаем целесообразным отметить основные этапы в эволюции представлений о природе света *).

Философы древности, размышлявшие о природе света, знали о зажигательных стеклах, о прямолинейном распространении света, о преломлении и отражении. Первые систематические описания оптических явлений, описания, о которых мы имеем некоторое представление, принадлежат греческим философам и математикам (Емпедоклу (490—430 гг. до н. э.), Евклиду (300 г. до н. э.)).

Из основателей новой философии следует отметить Рене Декарта (1596—1650 гг.), который сформулировал взгляды на природу света на основе метафизических представлений [8]. Декарт считал, что свет — это сжатие, распространяющееся в идеально упругой среде (эфире), которая заполняет все пространство, а различие цветов он объяснял вращательными движениями частиц этой среды с различными скоростями. Однако только после того, как Галилео Галилей (1564—1642 гг.), развивая механику, продемонстрировал мощь своего экспериментального метода, оптика получила прочную основу. Закон отражения был известен еще грекам; закон же преломления света был экспериментально установлен в 1621 г. Веллебродом Снеллиусом **) (1591—1626 гг.). В 1657 г. Пьер Ферма (1601—1665 гг.) выдвинул свой знаменитый «принцип наименьшего времени» ***) в следующей форме: «Природа всегда следует наикратчайшему пути». В соответствии с этим принципом свет распространяется по пути, требующему наименьшего времени; отсюда, а также

*) Для более полного ознакомления с историей оптики см. [1—6]. Подробный исторический обзор оптических исследований вплоть до настоящего времени см. в книге Уиттекера [7]. Первый том этой книги послужил основным источником при написании настоящего исторического введения.

**) Снеллиус умер в 1626 г., не опубликовав своего открытия. Закон преломления был впервые описан в «Диоптрике» Декарта без ссылки на Снеллиуса, хотя, как считают, Декарт был знаком с его рукописью на эту тему.

***) В письме к Кюро де ля Шамбру. Оно опубликовано в [9].

из предположения о различиях в «сопротивлениях» разных сред вытекает закон преломления света. Принцип Ферма имеет огромное философское значение и в свое время породил множество споров, так как его истолкование не свободно от телеологических положений, чуждых естественным наукам.

Впервые явление интерференции, а именно возникновение разноцветной окраски тонких пленок (в настоящее время такая картина называется «кольцами Ньютона»), было независимо обнаружено Робертом Бойлем (1627—1691 гг.) [10] и Робертом Гуком (1635—1703 гг.) [11]. Гук установил также наличие света в области геометрической тени, т. е. «дифракцию» света, однако это явление было замечено ранее Франциском Мария Гримальди (1618—1663 гг.) [12]. Гук был первым исследователем, который считал, что свет «состоит» из быстрых колебаний, распространяющихся мгновенно или с очень большой скоростью на любые расстояния, и что каждое колебание в однородной среде порождает сферу, радиус которой постоянно растет со временем *). С помощью таких представлений Гук пытался объяснить явление преломления и дать интерпретацию цвета. Однако природа цвета была выяснена лишь в 1666 г., когда Исаак Ньютон (1642—1727 гг.) обнаружил [13], что белый цвет с помощью призмы можно разложить на отдельные цветовые компоненты и что для каждого чистого цвета характерна своя степень преломления. Трудности, возникавшие в волновой теории при попытках объяснить прямолинейное распространение света и явление поляризации (открытое Гюйгенсом [14]), казались Ньютону настолько серьезными, что побудили его развить корпускулярную теорию (или теорию истечения), согласно которой свет распространяется от излучающего тела в виде мельчайших частиц.

Ко времени опубликования теории цвета Ньютона еще не было известно, распространяется ли свет мгновенно или нет. Конечность скорости света была обнаружена в 1675 г. Олафом Рёмером (1644—1710 гг.) при наблюдениях за затмениями спутников Юпитера [15].

Волновая теория света, среди первых последователей которой, как мы видели, находился Гук, была существенно улучшена и расширена Христианом Гюйгенсом (1629—1695 гг.) [14]. Он выдвинул принцип, названный позднее его именем, согласно которому каждую точку «эфира», до которой дошло световое возмущение, можно рассматривать как центр нового возмущения, распространяющегося в виде сферической волны; эти вторичные волны комбинируются таким образом, что их огибающая определяет волновой фронт в любой последующий момент времени. С помощью этого принципа Гюйгенсу удалось вывести законы преломления и отражения света. Он также объяснил двойное лучепреломление в исландском шпате (открытое в 1669 г. Эразмом Бартолинусом (1625—1698 гг.)), предположив, что при прохождении света через кристалл возникает, кроме первичной сферической, вторичная эллипсоидальная волна. В процессе своего исследования Гюйгенс обнаружил чрезвычайно важное явление, а именно явление поляризации: он показал, что каждый из двух лучей, возникающих после прохождения света через кристалл исландского шпата, можно погасить, пропуская его через второй такой же кристалл и вращая последний относительно направления луча. Однако объяснить поляризацию удалось только Ньютону; он предположил, что лучи имеют «стороны», и именно признание подобной «поперечности» света казалось ему непреодолимым возражением против волновой теории, поскольку ученым в то время были известны только продольные волны (из изучения распространения звука).

Отрицание волновой теории таким авторитетом, как Ньютон, привело к полному ее забвению в течение почти столетия. Однако иногда появлялись

*) В ранних волновых теориях Гука и Гюйгенса рассматривались единичные «импульсы», а не волновые цуги с определенными длинами волн.

ее случайные защитники *), например великий математик Леонард Эйлер (1707—1783 гг.) [16].

Только в начале XIX века были сделаны важнейшие открытия, приведшие к полному признанию волновой теории. Первым шагом в этом направлении послужило объяснение интерференции, выдвинутое в 1801 г. Томасом Юнгом (1773—1829 гг.), а также цветов тонких пленок [17]. Однако, поскольку идеи Юнга были развиты в основном лишь качественно, они не получили общего признания.

Примерно в это же время Этьен Луи Малюс (1775—1812 гг.) [18] обнаружил поляризацию света при отражении. Вероятно, в один из вечеров 1808 г. он наблюдал через кристалл исландского шпата отражение Солнца в оконном стекле и обнаружил, что при вращении кристалла вокруг линии зрения относительные интенсивности двух изображений, возникающих благодаря двойному лучепреломлению, изменяются. Однако Малюс не пытался найти объяснение этого явления, считая, по-видимому, что существовавшие тогда теории не в состоянии дать его.

Тем временем в работах Пьера Симона де Лапласа (1749—1827 гг.) и Жана-Батиста Био (1774—1862 гг.) развивалась далее корпускулярная теория. Ее сторонники предложили считать объяснение явления дифракции достойным премии, учрежденной на 1818 г. Парижской Академией наук, надеясь, что исследования в этой области полностью подтвердят корпускулярную теорию. Однако их надежды не оправдались — несмотря на сильное сопротивление, премия была присуждена Августину Жаку Френелю (1788—1827 гг.), исследование которого [19] основывалось на волновой теории и явилось первым из серии работ, полностью развенчавших в течение нескольких лет корпускулярную теорию. Сущность его исследования состояла в синтезе идеи Гюйгенса о построении волнового фронта как огибающей сферических волн и принципа интерференции Юнга. Этого, как показал Френель, оказалось достаточно для объяснения не только «прямолинейности» распространения света, но и небольших отклонений от «прямолинейности», т. е. явления дифракции. Френель решил задачи о дифракции на крае, небольших отверстиях и экране; наиболее убедительным оказалось экспериментальное подтверждение Араго предсказания, выведенного Пуассоном из теории Френеля и состоявшего в том, что в центре тени от круглого диска должно находиться светлое пятно.

В том же году (1818 г.) Френель занялся весьма важной проблемой влияния движения Земли на распространение света, а именно попытался выяснить, существует ли какое-нибудь различие между светом от звезд и светом от земных источников. Доменик Франсуа Араго (1786—1853 гг.) экспериментально обнаружил, что (помимо аберрации) никакого различия нет. На основании этих наблюдений Френель создал теорию о частичном увлечении светового эфира движущимися телами, теорию, которая была подтверждена в 1851 г. прямыми измерениями Армандом Ипполитом Луи Физо (1819—1896 гг.). Вместе с Араго Френель исследовал интерференцию поляризованных лучей света и обнаружил (в 1816 г.), что лучи, поляризованные во взаимно перпендикулярных плоскостях, никогда не интерферируют. Этот факт нельзя было согласовать с общепринятым тогда предположением о продольности световых волн. Юнг, узнавший об этом открытии от Араго, нашел в 1817 г. разгадку возникшего противоречия, предположив, что световые колебания поперечны.

Френель, сразу же оценив всю важность такого предположения, попытался подтвердить его, исходя из более надежной динамической основы [20], и вывел из него много следствий. Так, поскольку в жидкости могут существовать

) Последователем волновой теории света был и М. В. Ломоносов, который в работе, опубликованной в 1756 г., отстаивал представление о свете как о колебаниях эфира [62]. (*Прим. ред.*)

только продольные волны, эфир должен вести себя, как твердое тело; однако в то время теория упругих волн в твердых телах еще не была сформулирована. Вместо создания такой теории и вывода из нее следствий для оптики Френель «обернул» задачу и попытался выяснить свойства светового эфира из наблюдений. Он начал с изучения необычных законов распространения света в кристаллах; выяснение этих законов и сведение их к нескольким простым предположениям о природе элементарных волн является одним из важнейших достижений естественной науки. В 1832 г. Вильям Гамильтон (1805—1865 гг.) [21], сам внесший большой вклад в развитие оптики, обратил внимание на важное следствие, вытекающее из теории Френеля, а именно на возможность существования конической рефракции, которая была экспериментально обнаружена вскоре после этого Хамфри Ллойдом (1800—1881 гг.) [22].

Френель также первый сделал предположение (1821 г.), развитое позднее Коши, что для выяснения причины дисперсии необходимо учитывать молекулярную структуру вещества [23].

Динамические модели механизма колебаний эфира привели Френеля к законам (носящим теперь его имя), которые дают интенсивность и поляризацию световых лучей после преломления и отражения [24].

Работа Френеля столь надежно обосновала волновую теорию, что казалось совершенно излишним проведение контрольного эксперимента, впервые предложенного Араго, который был осуществлен в 1850 г. Фуко [25] и Физо и Бреже [26].

Корпускулярная теория объясняет преломление как притяжение световых частиц на границе двух сред оптически более плотной средой, откуда вытекает, что скорость света в более плотной среде больше; волновая же теория, согласно Гюйгенсу, дает меньшую скорость света в оптически более плотной среде. Непосредственное измерение скорости света в воздухе и воде полностью подтвердило вывод волновой теории *).

В последующие десятилетия была развита теория упругого эфира. Первым шагом в этом направлении явилось создание теории упругости для твердых тел. Она была сформулирована Клодом Луи Мария Анри Навье (1785—1836 гг.) [27], предположившим, что вещество состоит из бесчисленного количества частиц (точечных масс, атомов), взаимодействующих друг с другом вдоль линий, соединяющих пары частиц. Августину Луи Коши (1789—1857 гг.) [28] принадлежит привычный теперь вывод уравнений упругости для сплошной среды. Из других ученых, принимавших участие в развитии теории оптики, следует упомянуть Симона Дени Пуассона (1781—1840 гг.) [29], Джорджа Грина (1793—1841 гг.) [30], Джеймса Маккалаха (1809—1847 гг.) [31] и Франца Неймана (1798—1895 гг.) [32]. В настоящее время нет смысла подробно обсуждать ни сами теории, ни те трудности, с которыми они сталкивались, поскольку все эти трудности вытекали из требования (от него уже давно отказались), чтобы оптические явления можно было описывать в рамках механики. Достаточно упомянуть следующее. Рассмотрим две соприкасающиеся упругие среды и предположим, что в первой среде в направлении границы раздела распространяется поперечная волна. Согласно законам механики эта волна расщепится во второй среде на две: продольную и поперечную. Однако из экспериментов Араго и Френеля следует, что продольной упругой волны не должно быть. Разрешить это противоречие невозможно, не нарушая законов механики, содержащихся в граничных условиях для напряжений и натяжений. Различные теории, предложенные упомянутыми выше авторами, отличались друг от друга предполагаемыми граничными условиями, которые всегда оказывались несовместимыми с законами механики.

) По существу корпускулярная и волновая теории оперировали с разными понятиями скорости. Этот вопрос подробно разобран в [63]. (*Прим. ред.*)

Если считать эфир упругой твердой субстанцией, то как тогда ответить на следующий вопрос: каким образом движутся через такую среду планеты с огромными скоростями, не испытывая при этом никакого сопротивления? Джордж Габриэль Стокс (1819—1903 гг.) считал, что эту трудность можно обойти, если предположить, что скорости планет малы по сравнению со скоростями колебаний частиц эфира, образующих свет; например, известно, что такие твердые тела, как смола или сургуч, могут колебаться с большой частотой, но легко деформируются, если велик период колебания воздействующей силы. Подобные возражения кажутся нам теперь искусственными, поскольку мы не считаем необходимым, чтобы все явления природы можно было истолковать в рамках механических представлений.

Первый шаг на пути отхода от теории упругого эфира был сделан Маккалахом [33], постулировавшим существование среды со свойствами, которыми обычные тела не обладают. Последние накапливают энергию при деформации элементов объема, при вращении же накопления энергии не происходит. В эфире Маккалаха имеет место обратная ситуация. Законы распространений волн в такой среде весьма сходны с законами, вытекающими из уравнения Максвелла (для электромагнитных волн), которые являются основой современной оптики.

Несмотря на множество трудностей, теория упругого эфира доминировала в течение длительного времени, и многие выдающиеся физики XIX века внесли свой вклад в ее развитие. Кроме уже отмеченных ученых, необходимо упомянуть Вильяма Томсона (лорд Кельвин, 1824—1908 гг.) [34], Карла Неймана (1832—1925 гг.) [35], Джона Вильяма Стрэтта (лорд Рэлей, 1842—1919 гг.) [36] и Густава Кирхгофа (1824—1887 гг.) [37]. За это время были решены многие оптические проблемы, однако объяснение основ оптики оставалось неудовлетворительным.

В это же время практически независимо от оптических работ проводились исследования по электричеству и магнетизму, увенчавшиеся открытиями Майкла Фарадея (1791—1867 гг.) [38]. Джеймсу Кларку Максвеллу (1831—1879 гг.) [39] удалось подытожить все имевшиеся знания в этой области, сформулировав систему уравнений; наиболее важным их следствием оказалась возможность существования электромагнитных волн, распространяющихся со скоростью, величину которой можно вычислить из результатов чисто электрических измерений. Когда Рудольф Кольрауш (1809—1858 гг.) и Вильгельм Вебер (1804—1891) [40] выполнили эти измерения, скорость электромагнитных волн совпала со скоростью света. Отсюда Максвелл заключил, что свет представляет собой электромагнитные волны; его заключение было экспериментально подтверждено в 1888 г. Генрихом Герцем (1857—1894) [41]. Несмотря на это, электромагнитная теория Максвелла выдержала длительную борьбу, прежде чем получила всеобщее признание. По-видимому, одно из характерных свойств мышления человека состоит в том, что оно крайне неохотно отказывается от привычных представлений, особенно если приходится жертвовать ради этого конкретной картиной явления. В течение длительного времени сам Максвелл и его последователи пытались описать электромагнитное поле с помощью механических моделей. Только потом, когда идеи Максвелла стали более привычными, ученые постепенно оставили попытки «объяснения» его уравнений на основе механики; в настоящее время не возникает трудностей при представлении электромагнитного поля Максвелла в виде некой субстанции, не сводящейся ни к чему более простому.

Но даже электромагнитная теория света достигла со временем границ, за которыми она становится неприменимой. Она способна объяснить в общих чертах все явления, связанные с распространением света. Однако она не смогла описать процессы излучения и поглощения, которые определяются более тонкими особенностями взаимодействия вещества с оптическим полем.

Законы, управляющие этими процессами, являются предметом исследования современной оптики, даже более того,— современной физики. Их история начинается с открытия некоторых закономерностей в спектрах. Первым было открытие (в 1814—1817 гг.) темных линий в солнечном спектре Джозефом Фраунгофером (1787—1826 гг.) [42], названных его именем *), и их интерпретация как линий поглощения, данная в 1861 г. на основе экспериментов Робертом Вильгельмом Бунзеном (1811—1899 гг.) и Густавом Кирхгофом (1824—1887 гг.) [44]. Солнечный свет, обладающий непрерывным спектром, проходя через более холодные газы солнечной атмосферы, поглощается в атмосфере именно на тех длинах волн, которые излучают сами газы. Это открытие положило начало развитию спектрального анализа, в основе которого лежит утверждение, что все газообразные химические элементы обладают характерным линейчатым спектром. Изучение этих спектров было и остается главной задачей физических исследований; поскольку в таких экспериментах свет является предметом исследования и используются оптические методы, спектральный анализ рассматривается иногда как часть оптики. Однако вопрос об излучении и поглощении света атомами относится не к одной только оптике, так как в него входит и механика самого атома; спектральные закономерности раскрывают не столько природу света, сколько структуру излучающих частиц. Таким образом, спектроскопия из части оптики постепенно превратилась в самостоятельную дисциплину, дающую экспериментальное обоснование атомной и молекулярной физике. Эти вопросы, однако, выходят за рамки настоящей книги.

Что касается методов, то со временем стало ясно, что классическая механика не в состоянии дать правильное описание явлений, происходящих внутри атомов, и ее следует заменить квантовой теорией, история которой началась в 1900 г. с работы Макса Планка (1858—1947 гг.) [45]. Применив эту теорию к атому, Нильсу Бору (род. в 1885 г.) **) [46] удалось в 1913 г. объяснить простые закономерности в линейчатых спектрах газов. На основе этих первых работ и все возрастающего количества экспериментальных данных развилась современная квантовая механика (Гейзенберг, Борн, Иордан, де Бройль, Шредингер, Дирак) [47—52]. С ее помощью удалось существенно увеличить наши познания о структуре атомов и молекул.

Квантовая теория оказала сильное влияние на наши представления о природе света. Даже в своей первоначальной форме (в теории Планка) квантовая теория содержала предположение, полностью противоречащее классической физике, а именно предположение, что колебательная электрическая система передает свою энергию электромагнитному полю не непрерывно, а лишь конечными порциями, или «квантами» $\varepsilon = h\nu$, величины которых пропорциональны частоте света ν, а $h = 6{,}55 \cdot 10^{-27}$ *эрг/сек* — постоянная Планка***). Можно утверждать, что наличие постоянной h отличает современную физику от старой.

Прошло довольно много времени, прежде чем физики полностью осознали парадоксальный, почти иррациональный смысл уравнения Планка $\varepsilon = h\nu$. Заслуга в этом принадлежит в основном Эйнштейну и Бору. В 1905 г. Эйнштейн (1879—1955 гг.) на основании теории Планка возродил в новой форме корпускулярную теорию света [53], предположив, что планковские кванты энергии существуют в виде реальных частиц света, названных им *световыми квантами*, или *фотонами*. Таким образом, ему удалось объяснить некоторые явления, открытые ранее в связи с превращением энергии света в энергию

*) Волластон (1766—1828 гг.) наблюдал эти линии в 1802 г. [43], однако он не оценил свое открытие и неверно их интерпретировал.

**) Бор умер в 1962 г. (*Прим. ред.*)

***) В настоящее время считается, что $h = 6{,}62491 \cdot 10^{-27}$ *эрг/сек*. (*Прим. ред.*)

частиц и необъяснимые с помощью волновой теории. Главными среди них являются так называемый фотоэлектрический эффект и фотохимические эффекты. В явлениях такого рода не происходит передачи отдельной частице энергии, пропорциональной интенсивности света, как этого требует волновая теория, а свет скорее напоминает мелкий град. Энергия, сообщенная вторичным частицам, не зависит от интенсивности света, а определяется лишь его частотой (в соответствии с законом $\varepsilon=h\nu$). Из года в год росло число наблюдений, подтверждавших это свойство света, в результате чего сложилась ситуация, когда пришлось признать справедливость как волновой, так и корпускулярной теории света, причем первая подтверждалась явлением интерференции, а вторая — фотоэлектрическим эффектом. Только в самое последнее время развитие квантовой механики позволило частично объяснить такое парадоксальное положение, причем для этого пришлось отказаться от основного принципа старой физики, а именно, от принципа детерминированной причинности.

Развитие теории взаимодействия поля с веществом потребовало расширения методов квантовой механики (квантование полей). Для электромагнитного поля квантование было впервые проведено Дираком [54] и его работы составляют основу квантовой оптики. Обсуждение этих вопросов выходит, однако, за рамки настоящей книги.

Оптика движущихся тел является другой областью оптики, не затронутой в настоящей книге. Как и квантовая теория, она превратилась в широкий независимый раздел знания. Первым наблюденным явлением в этой области, отмеченным в 1728 г. Джеймсом Брэдли (1692—1762 гг.) [55], было явление аберрации «неподвижных звезд», т. е. обнаружение небольшого различия их угловых положений, связанного с движением Земли относительно направления светового луча. Брэдли правильно понял это явление, связав его с конечностью скорости распространения света, в результате чего ему удалось определить последнюю. Мы уже упоминали и другие явления, относящиеся к оптике движущихся сред: Френель первый заинтересовался увлечением света движущимися телами и показал, что световой эфир участвует в движении со скоростью, которая меньше скорости движущихся тел; затем Физо экспериментально продемонстрировал такое частичное увлечение света в опытах с текущей водой. Христиан Допплер (1803—1853 гг.) [56] исследовал эффекты, связанные с движением источника света или наблюдателя, и сформулировал хорошо известный принцип, названный его именем. До тех пор, пока теория упругого светового эфира считалась верной, а область исследований и точность измерений были достаточно ограниченными, идея Френеля о частичном увлечении света была способна объяснить все наблюдаемые явления. Электромагнитная же теория света встретилась здесь с трудностями фундаментального характера. Герц первый попытался обобщить уравнения Максвелла на случай движущихся тел. Однако его формулы противоречили некоторым электромагнитным и оптическим измерениям. Огромную роль сыграла теория Гендрика Антона Лоренца (1853—1928 гг.), который предположил, что «эфир в состоянии абсолютного покоя» является носителем электромагнитного поля, и вывел свойства материальных тел из взаимодействия элементарных электрических частиц — электронов. Ему удалось показать, что френелевские коэффициенты увлечения света можно получить из его теории и все известные в то время (1895 г.) явления можно объяснить на основании его гипотезы [57]. Однако в результате колоссального увеличения точности измерения оптических путей, достигнутого с помощью интерферометра Альберта Абрагама Майкельсона (1852—1931 гг.), возникла новая трудность: оказалось невозможным обнаружить «эфирный ветер», наличие которого следовало из теории «неподвижного эфира» [58, 59]. Эта трудность была преодолена в 1905 г. Альбертом Эйнштейном [60] в его специальной теории относительности.

Эта теория основана на критике старых представлений о времени и пространстве и приводит к отказу от евклидовой геометрии и обычного понятия одновременности. Ее дальнейшее развитие, так называемая общая теория относительности [61], характеризуется совершенно новым подходом к гравитационным явлениям — «геометризацией» пространственно-временного континуума.

Применение этой теории требует использования специальных математических и физических методов, которые (хотя они и связаны во многих случаях с оптикой) лучше рассматривать отдельно. Число же оптических явлений, в которых движение тел (т. е. источников света) играет существенную роль, весьма незначительно.

ЛИТЕРАТУРА *)

1. J. Priestley, History and Present state of Discoveries relating to Vision, Light and Colours, 2 Vols, London, 1772.
2. T. Young, A Course of Lectures on Natural Philosophy and the Mechanical Arts, London, 1845, Vol. 1, pp. 374—385.
3. E. Wilde, Geschichte der Optik vom Ursprung dieser Wissenschaft bis auf die gegenwärtige Zeit, 2 Vols, Berlin, 1838, 1843.
4. E. Mach, The Principles of Physical Optics, A historical and philosophical treatment 1926 (перевод с немецкого; первое немецкое издание 1913 г.).
5. E. Hoppe, Geschichte der Optik, Weber, Leipzig, 1926.
6. V. Ronchi, Storia della Luce, Zanichelli, Bologna, 2nd ed., 1952.
7. E. T. Whittaker, A History of the Theories of Aether and Electricity, Vol. I (The Classical Theory), 1952, Vol. II (The Modern Theories 1900—1926), 1953, T. Nelson a. Sons, London a. Edinburgh.
8. R. Descartes, Dioptrique, Météores, Leyden, 1637; Principia Philosophiae, Amsterdam, 1644.
9. P. Fermat, Oeuvres, Paris, 1891, Vol. 2, p. 354.
10. R. Boyle, The Philosophical Works, 2nd ed., London, 1738, Vol. II, p. 70.
11. R. Hooke, Micrographia, 1665, p. 47.
12. F. M. Grimaldi, Physico-Mathesis de lumine, coloribus, et iride, Bologna, 1665.
13. I. Newton, Phil. Trans, № 80, 3075 (1672).
14. C. Huygens, Traité de la lumière (работа закончена в 1678 г. и опубликована в Лейдене в 1690 г.) (X. Гюйгенс, Трактат о свете, ОНТИ, 1935).
15. O. Römer, Mém. Acad. Sci. Paris 10, 575 (1666—1669); J. de Sav., 223 (1676).
16. L. Euleri, Opuscula varii argumenti, Berlin, 1746, p. 169.
17. T. Young, Phil. Trans. Roy. Soc. xcii, 12, 387 (1802); Young's Works, Vol. 1, p. 202.
18. E. L. Malus, Nouveau Bull. d. Sci., par la Soc. Philomatique 1, 266 (1809); Mém. de la Soc. d'Arcueil 2 (1809).
19. A. Fresnel, Ann. Chim. et Phys. 1, 239 (1816); Oeuvres, Vol. 1, pp. 89, 129.
20. A. Fresnel, Oeuvres, Vol. 2, pp. 261, 479.
21. W. R. Hamilton, Trans. Roy. Irish Acad. 17, 1 (1833); Mathematical Papers, Vol. 1, Geometrical Optics, ed., by J. L. Synge a. W. Conway, Cambr. Univ. Press, 1931, p. 285.
22. H. Lloyd, Trans. Roy. Irish Acad. 17, 145 (1833).
23. A. Fresnel, Oeuvres, Vol. 2, p. 438.
24. A. Fresnel, Mém. de l'Acad. 11, 393 (1832); Oeuvres, Vol. 1, p. 767.
25. L. Foucault, C. R. Acad. Sci. 30, 551 (1850).
26. H. Fizeau, L. Breguet, C. R. Acad. Sci. 30, 562, 771 (1850).
27. C. L. M. H. Navier, Mém. de l'Acad. 7, 375 (представлена в 1821 г., опубликована в 1827 г.).
28. A. L. Cauchy, Exercise de Mathématiques 3, 160 (1828).
29. S. D. Poisson, Mém. de l'Acad. 8, 623 (1828).
30. G. Green, Trans. Cambr. Phil. Soc. (1838); Math. Papers, 245.
31. J. MacCullagh, Phil. Mag. 10, 42, 382 (1837); Proc. Roy. Irish. Acad. 18 (1837).
32. F. Neumann, Abh. Berl. Akad., Math. Kl. 1 (1835).
33. J. MacCullagh, Trans. Roy. Irish Acad. 21; Coll. Works, Dublin, 1880, p. 145.
34. W. Thomson, Phil. Mag. 26, 414 (1888); Baltimore Lectures, London, 1904.
35. C. Neumann, Math. Ann. 1, 325 (1869); 2, 182 (1870).
36. J. W. Strutt (Lord Rayleigh), Phil. Mag. 41, 519 (1871); 42, 81 (1871).
37. G. Kirchhoff, Berl. Abh. Physik, Abteilg 2, 57 (1876); Ges. Abh. 352; Berl. Ber. 641 (1882); Pogg. Ann. Physik u. Chem. (2) 18, 663 (1883); Ges. Abh., Nachtrag 22.

*) Здесь и далее литература, добавленная при переводе, отмечена звездочкой. (*Прим. ред.*)

38. M. Faraday, Experimental Researches in Electricity, London, 1839. (М. Фарадей, Избранные работы по электричеству, Гостехиздат, 1939.)
39. J. C. Maxwell, A Treatise on Electricity and Magnetism, 2 Vols, Oxford, 1873. (Дж. Максвелл, Избранные труды по теории электромагнитного поля, Гостехиздат, 1952.)
40. R. Kohlrausch, W. Weber, Pogg. Ann. Physik u. Chem. (2) **99**, 10 (1856).
41. H. Hertz, Sitzb. Berl. Akad. Wiss. Feb. 2, 1888; Wiedem. Ann. **34**, 551 (1888).
42. J. Fraunhofer, Gilberts Ann. **56**, 264 (1817).
43. W. H. Wollaston, Phil. Trans. Roy. Soc. 365 (1802).
44. R. Bunsen, G. Kirchhoff, Untersuchungen über das Sonnenspektrum und die Spektren der chemischen Elemente, Abb. Kgl. Akad. Wiss., Berlin, 1861, 1863.
45. M. Planck, Verh. d. deutsch phys. Ges. **2**, 202, 237 (1900). Ann. d. Physik **4**, 553 (1901).
46. N. Bohr, Phil. Mag. **26**, 1, 476, 857 (1913).
47. W. Heisenberg, Z.f. Phys. **33**, 879 (1925).
48. M. Born, P. Jordan, Z.f. Phis. **34**, 858 (1925).
49. M. Born, W. Heisenberg, P. Jordan, Z.f. Phys. **35**, 557 (1926).
50. L. de Broglie, Thèse, Paris, 1924; Ann. de Physique (10) **3**, 22 (1925).
51. E. Schrödinger, Ann. d. Physik **79**, 361, 489, 734 (1926); **80**, 437 (1926); **81**, 109 (1926).
52. P. A. M. Dirac, Proc. Roy. Soc. **A109**, 642 (1925); **A110**, 561 (1926).
53. A. Einstein, Ann. d. Physik **17**, 132 (1905); **20**, 199 (1906).
54. P. A. M. Dirac, Proc. Roy. Soc.**A114**, 243, 710 (1927).
55. J. Bradley, Phil. Trans, **35**, 637 (1728).
56. C. Doppler, Abh. Köngl. böhm. Geselsch. **2**, 466 (1842); Pogg. Ann. **68**, 1 (1847).
57. H. A. Lorentz, Versuch einer Theorie der electrischen und optischen Erscheinungen in bewegten Körpern, E. J. Brill, Leiden, 1895.
58. A. A. Michelson, Amer. J. Sci. (3) **22**, 20 (1881).
59. A. A. Michelson, E. W. Morley, Amer. J. Sci. (3) **34**, 333 (1887); Phil. Mag. **24**, 449 (1887).
60. A. Einstein, Ann. d. Physik **17**, 891 (1905).
61. A. Einstein, Berl. Sitz. 778, 799, 831, 844 (1915); Ann. d. Physik **49**, 769 (1916).
62*. М. В. Ломоносов, Слово о происхождении света, новую теорию о цветах представляющее. Сочинения М. В. Ломоносова, изд. Академии наук, 1898 г., т. IV, стр. 392.
63*. Г. С. Ландсберг, Оптика, Гостехиздат, 1957.

ГЛАВА 1

ОСНОВНЫЕ СВОЙСТВА ЭЛЕКТРОМАГНИТНОГО ПОЛЯ

§ 1.1. Электромагнитное поле

1.1.1. Уравнения Максвелла. При наличии электрических зарядов в пространстве устанавливается возбужденное состояние, которое называют *электромагнитным полем*. Его представляют двумя векторами **E** и **B**, именуемыми соответственно *электрическим вектором* и *магнитной индукцией*.

В элементарных курсах основными векторами поля по историческим причинам обычно считают векторы **E** и **H**, а посредством векторов **D** и **B** описывают влияние среды. Однако в общей теории по причинам, выясняющимся при изучении электродинамики движущихся сред, обязательно представление, указанное в тексте.

Уравнения Максвелла (1)—(4) (см. ниже) можно разделить на две группы: одна, содержащая **E** и **B**, состоит из двух однородных уравнений (с правой частью, равной нулю, а другая, содержащая **D** и **H**, — из двух неоднородных уравнении (с правой частью, отличной от нуля). При релятивистском преобразовании пространства и времени (преобразование Лорентца) уравнения каждой группы преобразуются совместно. При этом они остаются инвариантными, если величины $\mathbf{j}/c$ и ρ преобразуются как четырехмерный вектор, а каждая из пар **E**, **B** и **D**, **H** — как шестимерный вектор (антисимметричный тензор второго ранга). Поскольку в группу, состоящую из неоднородных уравнений, входят заряды и токи (которые представляют влияние среды), мы должны считать, что соответствующая пара (**D**, **H**) отражает влияние среды. Однако обычно за *вектор магнитного поля* принимают вектор **H**, а не **B**. Мы будем придерживаться той же терминологии, когда это не может привести к путанице.

Для того чтобы описать влияние поля на материальные объекты, необходимо ввести вторую группу векторов, а именно *плотность электрического тока* **j**, *электрическое смещение* **D** и *магнитный вектор* **H**.

Пространственные и временные производные пяти указанных векторов связаны *уравнениями Максвелла*. Эти уравнения, выполняющиеся в каждой точке, вблизи которой физические свойства среды непрерывны, имеют вид *)

$$\operatorname{rot}\mathbf{H}-\frac{1}{c}\dot{\mathbf{D}}=\frac{4\pi}{c}\mathbf{j}, \qquad (1)$$

$$\operatorname{rot}\mathbf{E}+\frac{1}{c}\dot{\mathbf{B}}=0. \qquad (2)$$

Точка над буквой означает дифференцирование по времени.

Векторные уравнения дополняются двумя скалярными соотношениями

$$\operatorname{div}\mathbf{D}=4\pi\rho, \qquad (3)$$

$$\operatorname{div}\mathbf{B}=0. \qquad (4)$$

Можно считать уравнение (3) определением плотности электрического заряда ρ, а соотношение (4) означает, что не существует свободных магнитных полюсов.

Из уравнения (1) следует (поскольку $\operatorname{div}\operatorname{rot}\equiv 0$), что

$$\operatorname{div}\mathbf{j}=-\frac{1}{4\pi}\operatorname{div}\dot{\mathbf{D}},$$

---

*) Здесь используется так называемая гауссова система единиц, т. е. электрические величины (**E**, **D**, **j** и ρ) измеряются в электростатических единицах, а магнитные величины (**H** и **B**) в электромагнитных единицах. Постоянная c в (1) и (2) связывает единицы заряда в обеих системах; она равна скорости света в вакууме и составляет приблизительно $3\cdot 10^{10}$ *см/сек* (более точное значение приведено в § 1.2). [В отличие от системы единиц, используемой в данной книге, в СССР принята Международная система единиц, сокращенно обозначаемая СИ [44*] *(Прим. ред.)*.]

или, используя (3)

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \tag{5}$$

По аналогии с соотношением, встречающимся в гидродинамике, уравнение (5) называется *уравнением непрерывности*. Оно отражает сохранение заряда в окрестности любой точки. Действительно, если мы проинтегрируем (5) по некоторой области пространства, то получим, используя теорему Гаусса,

$$\frac{d}{dt}\int \rho\, dV + \int \mathbf{j}\cdot\mathbf{n}\, dS = 0, \tag{6}$$

причем первый интеграл берется по объему этой области, а второй по поверхности, ограничивающей ее. Здесь $\mathbf{n}$ — единичный вектор внешней нормали. Из этого уравнения следует, что полный заряд

$$e = \int \rho\, dV, \tag{7}$$

заключенный внутри рассматриваемой области, может увеличиться только при наличии электрического тока

$$\mathcal{J} = \int \mathbf{j}\cdot\mathbf{n}\, dS. \tag{8}$$

Если все величины, характеризующие поле, не зависят от времени и отсутствуют токи ($\mathbf{j} = 0$), то поле называют *статическим*. Если эти величины не зависят от времени, но имеются токи ($\mathbf{j} \neq 0$), то мы говорим о *стационарном поле*. Векторы оптических полей очень быстро изменяются во времени, но источники полей обычно таковы, что при рассмотрении не мгновенных значений величин, а усредненных по любому макроскопическому интервалу времени, свойства поля оказываются не зависящими от времени. Часто термин «*стационарный*» употребляется в более широком смысле для описания поля указанного типа. В качестве примера укажем поле, образуемое непрерывным потоком излучения (скажем, от удаленной звезды) через оптическую систему.

1.1.2. Материальные уравнения. Уравнения Максвелла (1)—(4) связывают пять основных величин $\mathbf{E}$, $\mathbf{H}$, $\mathbf{B}$, $\mathbf{D}$ и $\mathbf{j}$. Для того чтобы при заданном распределении зарядов и токов уравнения допускали единственное решение для векторов поля, к этим уравнениям необходимо добавить соотношения, описывающие поведение веществ под влиянием поля. Такие соотношения называются *материальными уравнениями* *). В общем случае они довольно сложны, но для тел, находящихся в покое друг относительно друга (или в состоянии очень медленного движения) и состоящих из изотропных веществ (т. е. веществ, физические свойства которых в каждой точке не зависят от направления), эти уравнения принимают относительно простую форму **).

$$\mathbf{j} = \sigma \mathbf{E}, \tag{9}$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \tag{10}$$

$$\mathbf{B} = \mu \mathbf{H}. \tag{11}$$

Величина σ называется *удельной проводимостью*, ε — *диэлектрической проницаемостью*, а μ — *магнитной проницаемостью*.

Уравнение (9) является дифференциальной формой закона Ома. Вещества, для которых $\sigma \neq 0$ (вернее, не равно пренебрежимо малой величине; точный

---

*) Существует и другой способ описания поведения среды. Вместо величин $\varepsilon = D/E$ и $\mu = B/H$ можно рассматривать разности $\mathbf{D}-\mathbf{E}$ и $\mathbf{B}-\mathbf{H}$. Они имеют более простой физический смысл и их мы рассмотрим в гл. 2.

**) Более общие соотношения, применимые к движущимся телам, изучаются в релятивистской теории. Из общей теории нам необходим лишь следующий результат: в случае движущихся зарядов к току проводимости $\sigma\mathbf{E}$ добавляется конвекционный ток $\rho\mathbf{v}$, где $\mathbf{v}$ — скорость движущихся зарядов, ρ — плотность зарядов (см. стр. 31).

смысл этого неравенства здесь обсуждать невозможно), называются *проводниками*.

Очень хорошими проводниками служат металлы, но существуют и другие классы хорошо проводящих веществ, такие, как ионные растворы в жидкостях, а также в твердых телах. Проводимость металлов уменьшается с повышением температуры. Для других классов веществ, называемых *полупроводниками* (например, германий), проводимость сильно увеличивается с ростом температуры.

Вещества, для которых σ пренебрежимо мала, называются *изоляторами* или *диэлектриками*. Их электрические и магнитные свойства полностью определяются величинами ε и μ. Для большинства веществ магнитная проницаемость μ практически равна единице. Если это не так, т. е. если μ заметно отличается от единицы, то мы называем такое вещество *магнетиком*. В частности, если $\mu > 1$, то вещество называют *парамагнетиком* (например, платина, кислород, азот), если же $\mu < 1$,— то *диамагнетиком* (например, висмут, медь, водород, вода).

Для чрезвычайно сильных полей, которые получаются, например, при фокусировке света, генерируемого оптическим мазером *), к правым частям материальных уравнений, возможно, придется добавить члены, содержащие компоненты векторов поля в степени, большей единицы **).

Во многих случаях величины σ, ε и μ не зависят от напряженности полей, однако часто поведение вещества невозможно описать таким простым способом. Например, ток в газе свободных ионов, который определяется средней скоростью ионов, в любой момент времени зависит не от мгновенного значения **E**, а от всех его предыдущих значений. В так называемых *ферромагнитных* веществах (вещества, являющиеся очень сильными магнетиками, например железо, кобальт и никель) магнитная индукция **B** определяется предысторией поля **H**, а не его мгновенным значением. В этом случае говорят, что вещество проявляет *гистерезис*. В некоторых диэлектриках наблюдается подобная же зависимость от предыстории для электрического смещения. К счастью, гистерезисные эффекты для высокочастотных полей, встречающихся в оптике, как правило, незначительны.

В основной части настоящей книги мы будем рассматривать распространение света в таких средах, в которых не происходит заметного его ослабления (например, воздух, стекло). Подобные среды называются *прозрачными*. Они должны быть электрически непроводящими ($\sigma = 0$), поскольку наличие проводимости приводит к выделению джоулева тепла (см. п. 1.1.4) и, следовательно, к потерям электромагнитной энергии. Оптические свойства проводящих сред рассматриваются в гл. 13.

1.1.3. Граничные условия на поверхностях раздела. Уравнения Максвелла были получены лишь для областей пространства, в которых физические свойства среды (характеризующиеся величинами ε и μ) непрерывны. В оптике часто встречается ситуация, когда эти свойства резко меняются на одной или нескольких поверхностях. Можно ожидать, что тогда векторы **E**, **H**, **B** и **D** также будут претерпевать разрыв, а величины ρ и **j** выродятся в соответствующие поверхностные величины. Выведем соотношения, описывающие переход через такую поверхность раздела.

---

*) В русской литературе мазеры в оптическом диапазоне называются лазерами. (*Прим. ред.*)

) Нелинейность соотношения между вектором смещения **D и электрическим полем **E** впервые была найдена указанным способом в работе [1]. Теория распространения, преломления и отражения электромагнитных волн при условиях, когда необходимо учитывать такие нелинейные эффекты, развита в работах [2, 3]. Подробное рассмотрение нелинейных эффектов проводится в работах [4]. [Нелинейные оптические эффекты, рассмотрены также в работах [45*, 46*] (*Прим. ред.*).]

Заменим поверхность резкого раздела T тонким переходным слоем, внутри которого ε и μ быстро, но непрерывно меняются от значений около T с одной стороны слоя до их значений вблизи T с другой его стороны. Внутри этого слоя построим небольшой квазицилиндр, ограниченный с боков «частоколом» нормалей к T. Основаниями цилиндра на каждой стороне T служат небольшие площадки δA_1 и δA_2, параллельные поверхности T (рис. 1.1). Поскольку во всем цилиндре вектор $\mathbf{B}$ и его производные непрерывны, мы можем применить теорему Гаусса к интегралу от div $\mathbf{B}$, взятому по объему цилиндра. Тогда, согласно (4), получим

$$\int \operatorname{div} \mathbf{B}\, dV = \int \mathbf{B}\cdot\mathbf{n}\, dS = 0. \tag{12}$$

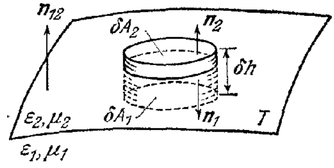


Рис. 1.1. К выводу граничных условий для нормальных компонент $\mathbf{B}$ и $\mathbf{D}$.

Здесь $\mathbf{n}$ — единичный вектор внешней нормали; второй интеграл берется по поверхности цилиндра.

Так как площадки δA_1 и δA_2 предполагаются малыми, можно считать, что на них $\mathbf{B}$ принимает постоянные значения $\mathbf{B}^{(1)}$ и $\mathbf{B}^{(2)}$. Тогда выражение (12) можно заменить следующим:

$$\mathbf{B}^{(1)}\cdot\mathbf{n}_1\delta A_1 + \mathbf{B}^{(2)}\cdot\mathbf{n}_2\delta A_2 + \text{вклад от стенок} = 0. \tag{13}$$

Если высота цилиндра δh стремится к нулю, переходный слой переходит в поверхность, а вклад от стенок цилиндра исчезает при условии, что отсутствует поверхностный поток магнитной индукции. Такой поток никогда не наблюдается и, следовательно, в пределе

$$(\mathbf{B}^{(1)}\cdot\mathbf{n}_1 + \mathbf{B}^{(2)}\cdot\mathbf{n}_2)\,\delta A = 0, \tag{14}$$

где δA — площадь пересечения нашего цилиндра с поверхностью T. Если $\mathbf{n}_{12}$— единичный вектор нормали, направленный из первой среды во вторую, то $\mathbf{n}_1 = -\mathbf{n}_{12}$, $\mathbf{n}_2 = \mathbf{n}_{12}$, и из соотношения (14) получим

$$\mathbf{n}_{12}\cdot(\mathbf{B}^{(2)} - \mathbf{B}^{(1)}) = 0, \tag{15}$$

т. е. *нормальная компонента вектора магнитной индукции непрерывна на поверхности раздела.*

Подобным же образом можно рассмотреть электрическое смещение $\mathbf{D}$, но в этом случае при наличии зарядов появится дополнительный член. Вместо (12) мы получим из (3)

$$\int \operatorname{div} \mathbf{D}\, dV = \int \mathbf{D}\cdot\mathbf{n}\, dS = 4\pi \int \rho\, dV. \tag{16}$$

При слиянии площадок δA_1 и δA_2 полный заряд остается конечным и, следовательно, объемная плотность становится бесконечной. При этом вместо объемной плотности заряда ρ необходимо ввести *поверхностную плотность заряда* $\hat{\rho}$, определяемую соотношением *)

$$\lim_{\delta h \to 0} \int \rho\, dV = \int \hat{\rho}\, dA. \tag{17}$$

*) В дальнейшем нам понадобится выразить поверхностную плотность заряда и поверхностную плотность тока через дельта-функцию Дирака (см. приложение 4). Если уравнение поверхности раздела имеет вид $F(x, y, z) = 0$, то

$$\rho = \hat{\rho}\,|\operatorname{grad} F|\,\delta(F), \tag{17a}$$

$$\mathbf{j} = \hat{\mathbf{j}}\,|\operatorname{grad} F|\,\delta(F). \tag{18a}$$

Можно легко проверить правильность этих выражений, подставив их в (17) и (18) и использовав соотношение $dF = |\operatorname{grad} F|\,dh$ и фильтрующее свойство дельта-функции Дирака.

Позже нам понадобится также понятие *поверхностной плотности тока* $\hat{\mathbf{j}}$, которая определяется аналогичным образом, а именно

$$\lim_{\delta h \to 0} \int \mathbf{j}\, dV = \int \hat{\mathbf{j}}\, dA. \qquad (18)$$

Если площадку δA и высоту δh выбрать достаточно малыми, то из (16) получим

$$\mathbf{D}^{(1)} \cdot \mathbf{n}_1 \delta A_1 + \mathbf{D}^{(2)} \cdot \mathbf{n}_2 \delta A + \text{вклад от стенок} = 4\pi \hat{\rho} \delta A.$$

Вклад от стенок стремится к нулю с уменьшением δh, и поэтому в пределе при $\delta h \to 0$ получим

$$\mathbf{n}_{12} \cdot (\mathbf{D}^{(2)} - \mathbf{D}^{(1)}) = 4\pi \hat{\rho}, \qquad (19)$$

т. е. *при наличии на поверхности раздела слоя с поверхностной плотностью заряда* $\hat{\rho}$, *нормальная компонента вектора электрического смещения при переходе через эту поверхность испытывает скачок, равный* $4\pi\hat{\rho}$.

Исследуем теперь поведение тангенциальных компонент. Заменим поверхность резкого раздела переходным слоем, а цилиндр, показанный на рис. 1.1, «прямоугольной» площадкой, стороны которой параллельны и перпендикулярны поверхности T (рис. 1.2).

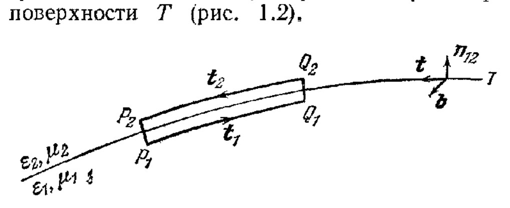


Рис. 1.2. К выводу граничных условий для тангенциальных компонент **E** и **H**.

Пусть **b** — единичный вектор, перпендикулярный плоскости рассматриваемого прямоугольника. Тогда на основании теоремы Стокса получим из (2)

$$\int \operatorname{rot} \mathbf{E} \cdot \mathbf{b}\, dS = \int \mathbf{E} \cdot d\mathbf{r} = -\frac{1}{c} \int \dot{\mathbf{B}} \cdot \mathbf{b}\, dS; \qquad (20)$$

здесь первый и третий интегралы берутся по площади прямоугольника, а второй — вдоль его границ. Если длины $P_1Q_1 (=\delta s_1)$ и $P_2Q_2 (=\delta s_2)$ малы, то на каждой из этих сторон вектор **E** можно заменить постоянными значениями $\mathbf{E}^{(1)}$ и $\mathbf{E}^{(2)}$. Подобным же образом постоянным значением можно заменить и вектор $\dot{\mathbf{B}}$. Тогда из (20) найдем

$$\mathbf{E}^{(1)} \cdot \mathbf{t}_1 \delta s_1 + \mathbf{E}^{(2)} \cdot \mathbf{t}_2 \delta s_2 + \text{вклад от концов} = -\frac{1}{c} \dot{\mathbf{B}} \cdot \mathbf{b} \delta s\, \delta h, \qquad (21)$$

где δs — линейный элемент, по которому прямоугольник пересекается с поверхностью раздела. Если теперь постепенно уменьшать высоту прямоугольника, то вклад от концов P_1P_2 и Q_1Q_2 будет стремиться к нулю при условии, что **E** в пределе не имеет достаточно резких особенностей. Такую возможность мы исключим. Предположим также, что остается конечным и $\dot{\mathbf{B}}$; тогда в пределе при $\delta h \to 0$ получим

$$(\mathbf{E}^{(1)} \cdot \mathbf{t}_1 + \mathbf{E}^{(2)} \cdot \mathbf{t}_2)\, \delta s = 0. \qquad (22)$$

Если **t** — единичный вектор касательной к поверхности (см. рис. 1.2), то $\mathbf{t}_1 = -\mathbf{t} = -\mathbf{b} \times \mathbf{n}_{12}$, $\mathbf{t}_2 = \mathbf{t} = \mathbf{b} \times \mathbf{n}_{12}$, и из (22) следует

$$\mathbf{b} \cdot [\mathbf{n}_{12} \times (\mathbf{E}^{(2)} - \mathbf{E}^{(1)})] = 0.$$

Так как ориентация прямоугольника, а следовательно, и единичного вектора $\mathbf{b}$ произвольна, ясно, что

$$\mathbf{n}_{12}\times(\mathbf{E}^{(2)}-\mathbf{E}^{(1)})=0, \tag{23}$$

т. е. *тангенциальная компонента электрического вектора непрерывна на поверхности раздела.*

Наконец, рассмотрим поведение тангенциальной компоненты магнитного вектора. Анализ проводится аналогичным образом, но при наличии токов появится дополнительный член. Вместо (21) в этом случае получим

$$\mathbf{H}^{(1)}\cdot\mathbf{t}_1\delta s_1+\mathbf{H}^{(2)}\cdot\mathbf{t}_2\delta s_2+\text{вклад от концов}=\frac{1}{c}\dot{\mathbf{D}}\cdot\mathbf{b}\,\delta s\,\delta h+\frac{4\pi}{c}\hat{\mathbf{j}}\cdot\mathbf{b}\,\delta s. \tag{24}$$

Переходя, как и раньше, к пределу $\delta h\to 0$, находим

$$\mathbf{n}_{12}\times(\mathbf{H}^{(2)}-\mathbf{H}^{(1)})=\frac{4\pi}{c}\hat{\mathbf{j}}. \tag{25}$$

Из (25) следует, что *при наличии тока с поверхностной плотностью* $\hat{\mathbf{j}}$ *тангенциальная компонента магнитного вектора (рассматриваемая как вектор) испытывает скачок, равный*

$$\frac{4\pi}{c}\hat{\mathbf{j}}\times\mathbf{n}_{12}.$$

Помимо разрывов, связанных с резким изменением физических свойств среды, векторы поля могут также претерпевать разрывы из-за присутствия источника, который начинает излучать в некоторый момент времени $t=t_0$. Тогда возмущение будет распространяться в окружающем пространстве и в любой последующий момент времени $t_1>t_0$ заполнит вполне определенную область. На (движущихся) границах этой области векторы поля будут резко изменяться от конечных значений до нуля за ее пределами.

Различные случаи разрыва можно описать уравнениями Максвелла в интегральной форме (см., например, [5], стр. 11 или [6], стр. 6). Общие условия разрывности можно также записать в форме уравнений в конечных разностях; вывод этих уравнений приведен в приложении 6.

1.1.4. Закон сохранения энергии для электромагнитного поля. В рамках электромагнитной теории интенсивность света интерпретируется как поток энергии поля. Поэтому необходимо вспомнить формулировку закона сохранения энергии в теории Максвелла. Из (1) и (2) следует, что

$$\mathbf{E}\cdot\operatorname{rot}\mathbf{H}-\mathbf{H}\cdot\operatorname{rot}\mathbf{E}=\frac{4\pi}{c}\mathbf{j}\cdot\mathbf{E}+\frac{1}{c}\mathbf{E}\cdot\dot{\mathbf{D}}+\frac{1}{c}\mathbf{H}\cdot\dot{\mathbf{B}}. \tag{26}$$

Члены, стоящие слева, с помощью хорошо известного векторного тождества можно выразить через дивергенцию векторного произведения $\mathbf{H}$ и $\mathbf{E}$, т. е.

$$\mathbf{E}\cdot\operatorname{rot}\mathbf{H}-\mathbf{H}\cdot\operatorname{rot}\mathbf{E}=-\operatorname{div}(\mathbf{E}\times\mathbf{H}). \tag{27}$$

Из (26) и (27) получим

$$\frac{1}{c}(\mathbf{E}\cdot\dot{\mathbf{D}}+\mathbf{H}\cdot\dot{\mathbf{B}})+\frac{4\pi}{c}\mathbf{j}\cdot\mathbf{E}+\operatorname{div}(\mathbf{E}\times\mathbf{H})=0. \tag{28}$$

Умножив это равенство на $c/4\pi$, проинтегрировав по произвольному объему и использовав теорему Гаусса, найдем

$$\frac{1}{4\pi}\int(\mathbf{E}\cdot\dot{\mathbf{D}}+\mathbf{H}\cdot\dot{\mathbf{B}})\,dV+\int\mathbf{j}\cdot\mathbf{E}\,dV+\frac{c}{4\pi}\int(\mathbf{E}\times\mathbf{H})\cdot\mathbf{n}\,dS=0. \tag{29}$$

Здесь последний интеграл берется по границе объема, а $\mathbf{n}$ — единичный вектор внешней нормали.

Соотношение (29) непосредственно вытекает из уравнения Максвелла и поэтому выполняется независимо от справедливости материальных уравнений (9)—(11). Как мы увидим, оно выражает закон сохранения энергии для

электромагнитного поля. Здесь мы рассмотрим его лишь для случая, когда удовлетворяются материальные уравнения (9)—(11). Позже (см. гл. 14) будет проведено обобщение этого закона для случая анизотропных сред, где материальные уравнения принимают более сложную форму.

Используя материальные уравнения, найдем

$$\left.\begin{aligned} \frac{1}{4\pi}(\mathbf{E}\cdot\dot{\mathbf{D}}) &= \frac{1}{4\pi}\mathbf{E}\cdot\frac{\partial}{\partial t}(\varepsilon\mathbf{E}) = \frac{1}{8\pi}\frac{\partial}{\partial t}(\varepsilon\mathbf{E}^2) = \frac{1}{8\pi}\frac{\partial}{\partial t}(\mathbf{E}\cdot\mathbf{D}), \\ \frac{1}{4\pi}(\mathbf{H}\cdot\dot{\mathbf{B}}) &= \frac{1}{4\pi}\mathbf{H}\cdot\frac{\partial}{\partial t}(\mu\mathbf{H}) = \frac{1}{8\pi}\frac{\partial}{\partial t}(\mu\mathbf{H}^2) = \frac{1}{8\pi}\frac{\partial}{\partial t}(\mathbf{H}\cdot\mathbf{B}). \end{aligned}\right\} \tag{30}$$

Полагая

$$w_e = \frac{1}{8\pi}\mathbf{E}\cdot\mathbf{D}, \qquad w_m = \frac{1}{8\pi}\mathbf{H}\cdot\mathbf{B} \tag{31}$$

и

$$W = \int (w_e + w_m)\,dV, \tag{32}$$

преобразуем соотношение (29) к виду

$$\frac{dW}{dt} + \int \mathbf{j}\cdot\mathbf{E}\,dV + \frac{c}{4\pi}\int (\mathbf{E}\times\mathbf{H})\cdot\mathbf{n}\,dS = 0. \tag{33}$$

Покажем, что величина W представляет полную энергию, заключенную внутри объема, и, следовательно, w_e можно отождествить с *плотностью электрической энергии*, а w_m — *с плотностью магнитной энергии* поля *).

Для отождествления W с полной энергией, нужно доказать, что для замкнутой системы (т. е. системы, в которой можно пренебречь полем на граничной поверхности) изменение W обусловливается работой, проделанной полем над материальными заряженными телами, входящими в систему. Достаточно показать это для медленного движения названных тел, причем мы вправе считать последние столь малыми, что можем рассматривать их как точечные заряды e_k ($k = 1, 2, \ldots$). Обозначим скорость заряда e_k через $\mathbf{v}_k$ ($v_k \ll c$).

Сила, действующая со стороны поля ($\mathbf{E}$, $\mathbf{B}$) на заряд, движущийся со скоростью $\mathbf{v}$, определяется так называемым *законом Лоренца*

$$\mathbf{F} = e\left(\mathbf{E} + \frac{1}{c}[\mathbf{v}\times\mathbf{B}]\right), \tag{34}$$

который основан на эксперименте. Отсюда следует, что если все заряды e_k за время δt смещаются на $\delta\mathbf{x}_k$ ($k = 1, 2, \ldots$), то полная проделанная работа равна

$$\delta A = \sum_k \mathbf{F}_k\cdot\delta\mathbf{x}_k = \sum_k e_k\left(\mathbf{E}_k + \frac{1}{c}[\mathbf{v}_k\times\mathbf{B}_k]\right)\cdot\delta\mathbf{x}_k = \sum_k e_k\mathbf{E}_k\cdot\delta\mathbf{x}_k = \sum_k e_k\mathbf{E}_k\cdot\mathbf{v}_k\,\delta t,$$

так как $\delta\mathbf{x}_k = \mathbf{v}_k\delta t$. Если число заряженных частиц велико, распределение заряда можно считать непрерывным. Введем плотность заряда ρ (т. е. полный заряд единицы объема); тогда последнее равенство примет вид

$$\delta A = \delta t \int \rho\mathbf{v}\cdot\mathbf{E}\,dV, \tag{35}$$

*) В общем случае эти величины определяются выражениями

$$w_e = \frac{1}{4\pi}\int \mathbf{E}\cdot d\mathbf{D}, \qquad w_m = \frac{1}{4\pi}\int \mathbf{H}\cdot d\mathbf{B}.$$

Они переходят в (31), когда соотношение между $\mathbf{E}$ и $\mathbf{D}$ и между $\mathbf{H}$ и $\mathbf{B}$ линейно, как предполагается здесь.

причем интегрирование проводится по произвольному объему. В уравнениях Максвелла не содержится в явном виде скорость $\mathbf{v}$, однако ее можно ввести, используя полученный Рентгеном [7] экспериментальный результат, согласно которому *конвекционный ток* (т. е. течение движущихся зарядов) создает такой же электромагнитный эффект, как и ток проводимости в проволоке. Следовательно, плотность тока $\mathbf{j}$, фигурирующую в уравнениях Максвелла, можно разделить на две части

$$\mathbf{j} = \mathbf{j}_c + \mathbf{j}_v, \tag{36}$$

где $\mathbf{j}_c = \sigma\mathbf{E}$ — плотность тока проводимости, а $\mathbf{j}_v = \rho\mathbf{v}$ — плотность конвекционного тока. Тогда выражение (35) запишется в виде

$$\delta A = \delta t \int \mathbf{j}_v \cdot \mathbf{E}\, dV. \tag{37}$$

Определим вектор $\mathbf{S}$ и скаляр Q соотношениями

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E}\times\mathbf{H}, \tag{38}$$

$$Q = \int \mathbf{j}_c \cdot \mathbf{E}\, dV = \int \sigma \mathbf{E}^2\, dV. \tag{39}$$

Теперь с помощью (35) и (36) найдем

$$\int \mathbf{j}\cdot\mathbf{E}\, dV = Q + \int \mathbf{j}_v \cdot \mathbf{E}\, dV = Q + \frac{\delta A}{\delta t}, \tag{40}$$

где вторая функция, конечно, не является полной производной от функции координат и времени. Уравнение (33) примет теперь вид

$$\frac{dW}{dt} = -\frac{\delta A}{\delta t} - Q - \int \mathbf{S}\cdot\mathbf{n}\, dS. \tag{41}$$

Для непроводящей среды ($\sigma = 0$) имеем $Q = 0$. Будем также считать граничную поверхность расположенной так далеко, что можно пренебречь полем на ней, вызванным электромагнитными процессами, происходящими внутри нее. Тогда $\int \mathbf{S}\cdot\mathbf{n}\, dS = 0$ и интегрирование выражения (41) дает

$$W + A = \text{const}. \tag{42}$$

Следовательно, для изолированной системы увеличение W в единицу времени обусловливается работой, проделанной над системой в течение этого времени. Полученный результат подтверждает наше определение электромагнитной энергии выражением (32).

Член Q (он называется *джоулевым теплом*) представляет собой диссипацию энергии в проводнике ($\sigma \neq 0$) из-за наличия сопротивления. Если поле достигает граничной поверхности, то, согласно (41), происходит дополнительное уменьшение энергии. Поэтому поверхностный интеграл должен определять поток энергии через эту граничную поверхность. Вектор $\mathbf{S}$ известен как *вектор Пойнтинга* и представляет количество энергии, протекающее за одну секунду через единичную площадку, перпендикулярную к направлениям $\mathbf{E}$ и $\mathbf{H}$.

Необходимо отметить, что интерпретация $\mathbf{S}$ как потока энергии (точнее, как плотности этого потока) вносит известную степень произвола. Согласно (41) физическим смыслом обладает не сама величина $\mathbf{S}$, а интеграл от $\mathbf{S}\cdot\mathbf{n}$, взятый по замкнутой поверхности. Ясно, что из значения интеграла нельзя вывести однозначное заключение о детальном распределении $\mathbf{S}$, поэтому возможны и другие определения плотности потока энергии. К $\mathbf{S}$ всегда можно прибавить ротор произвольного вектора, поскольку такой член не вносит вклада в поверхностный интеграл, что следует из теоремы Гаусса и тождества

$\operatorname{div}\operatorname{rot}\equiv 0$ *). Однако при осторожном применении этого определения, в частности для средних значений небольших, но конечных интервалов пространства или времени, никакого противоречия с экспериментом не возникает. Поэтому мы примем приведенное выше определение плотности потока энергии через вектор Пойнтинга.

Наконец, отметим, что в непроводящей среде ($\sigma = 0$) в отсутствие механической работы ($A = 0$) закон сохранения энергии можно записать в форме гидродинамического уравнения непрерывности для несжимаемой жидкости, а именно

$$\frac{\partial w}{\partial t} + \operatorname{div}\mathbf{S} = 0 \qquad (w = w_e + w_m). \tag{43}$$

Гидродинамическая модель часто оказывается полезной при описании распространения света, в частности в геометрической оптике или при рассмотрении скалярных дифракционных полей, так как она дает картину распространения энергии в простой графической форме. В оптике наибольший интерес представляет усредненный вектор Пойнтинга. Величина его служит мерой интенсивности света, а направление указывает направление распространения света.

§ 1.2. Волновое уравнение и скорость света

Уравнения Максвелла связывают между собой векторы поля системой дифференциальных уравнений. Дифференциальные уравнения, которым должен удовлетворять каждый из векторов в отдельности, можно получить путем исключения остальных векторов. Мы ограничимся рассмотрением области поля, не содержащей ни зарядов, ни токов, т. е. $\mathbf{j} = 0$ и $\rho = 0$.

Подставим выражение для $\mathbf{B}$ из материального уравнения (1.1.11) **) во второе уравнение Максвелла (1.1.2), разделим обе его части на μ и применим операцию rot. Это дает

$$\operatorname{rot}\left(\frac{1}{\mu}\operatorname{rot}\mathbf{E}\right) + \frac{1}{c}\operatorname{rot}\dot{\mathbf{H}} = 0. \tag{1}$$

Продифференцируем затем первое уравнение Максвелла (1.1.1) по времени, используем уравнение для $\mathbf{D}$ (1.1.10) и исключим $\operatorname{rot}\dot{\mathbf{H}}$ из системы двух уравнений, содержащей получающееся уравнение и уравнение (1). Тогда получим

$$\operatorname{rot}\left(\frac{1}{\mu}\operatorname{rot}\mathbf{E}\right) + \frac{\varepsilon}{c^2}\ddot{\mathbf{E}} = 0. \tag{2}$$

Если использовать тождества $\operatorname{rot} u\mathbf{v} \equiv u \operatorname{rot}\mathbf{v} + (\operatorname{grad} u) \times \mathbf{v}$ и $\operatorname{rot}\operatorname{rot} \equiv \operatorname{grad}\operatorname{div} - \nabla^2$, то (2) примет вид

$$\nabla^2\mathbf{E} - \frac{\varepsilon\mu}{c^2}\ddot{\mathbf{E}} + [\operatorname{grad}(\ln\mu)] \times \operatorname{rot}\mathbf{E} - \operatorname{grad}\operatorname{div}\mathbf{E} = 0. \tag{3}$$

Используя снова материальное уравнение для $\mathbf{D}$ и применяя тождество $\operatorname{div} u\mathbf{v} \equiv u \operatorname{div}\mathbf{v} + \mathbf{v}\cdot\operatorname{grad} u$, найдем из (1.1.3) соотношение

$$\varepsilon \operatorname{div}\mathbf{E} + \mathbf{E}\operatorname{grad}\varepsilon = 0. \tag{4}$$

Следовательно, уравнение (3) можно записать в виде

$$\nabla^2\mathbf{E} - \frac{\varepsilon\mu}{c^2}\ddot{\mathbf{E}} + [\operatorname{grad}(\ln\mu)] \times \operatorname{rot}\mathbf{E} + \operatorname{grad}[\mathbf{E}\cdot\operatorname{grad}(\ln\varepsilon)] = 0. \tag{5}$$

---

*) Современная теория поля допускает даже больший произвол, связанный с возможностью добавлять к лагранжиану поля четырехмерную дивергенцию. При этом допустимы другие выражения как для потока энергии, так и для плотности энергии.

**) Здесь и далее при ссылках на формулы, приведенные в других параграфах, мы будем указывать также номер главы и номер параграфа. Например, ссылка на формулу (5.3.13) означает, что имеется в виду формула (13), приведенная в § 3 гл. 5. (*Прим. ред.*)

Подобным же образом получается уравнение для **H**

$$\nabla^2\mathbf{H}-\frac{\varepsilon\mu}{c^2}\ddot{\mathbf{H}}+[\operatorname{grad}(\ln\varepsilon)]\times\operatorname{rot}\mathbf{H}+\operatorname{grad}[\mathbf{H}\cdot\operatorname{grad}(\ln\mu)]=0. \tag{6}$$

В частности, если среда однородна, то grad (ln ε) = grad (ln μ) = 0, и соотношения (5) и (6) принимают вид

$$\nabla^2\mathbf{E}-\frac{\varepsilon\mu}{c^2}\ddot{\mathbf{E}}=0,\qquad \nabla^2\mathbf{H}-\frac{\varepsilon\mu}{c^2}\ddot{\mathbf{H}}=0. \tag{7}$$

Это обычные уравнения волнового движения. Они означают, что существуют электромагнитные волны, распространяющиеся со скоростью *)

$$v=\frac{c}{\sqrt{\varepsilon\mu}}. \tag{8}$$

Впервые постоянная c была определена Р. Кольраушем и В. Вебером в 1856 г. из отношения значений емкости конденсатора, измеренных в электростатических и электромагнитных единицах. Оказалось, что она совпадает со скоростью света в вакууме. Используя этот результат, Максвелл развил свою электромагнитную теорию света, предсказывающую существование электромагнитных волн. Правильность его предсказания была подтверждена знаменитыми экспериментами Г. Герца (см. «Историческое введение»).

В любой волновой теории света элементарным процессом считают гармоническую волну в пространстве — времени (в простейшей форме она будет рассмотрена в §§ 1.3 и 1.4). Если ее частота лежит в интервале от $4\cdot10^{14}$ *сек*$^{-1}$ до $7{,}5\cdot10^{14}$ *сек*$^{-1}$ (приблизительно), то она вызывает у человека физиологическое ощущение определенного цвета. (Обратное, однако, неверно: окрашенный свет, вызывающий определенное субъективное цветовое ощущение, может быть совокупностью гармонических волн с весьма различными частотами.) Действительная связь между цветом и частотой очень сложна и не будет изучаться в настоящей книге **).

Первое определение скорости света было сделано Рёмером в 1675 г. на основе наблюдений затмений первого спутника Юпитера; позднее другим способом (путем наблюдения аберрации неподвижных звезд) ее измерил Брэдли (1728 г.) ***).

Первые измерения скорости света с земными источниками были выполнены Физо в 1849 г. При таких измерениях необходимо использовать модулятор, который разбивает пучок на части ****). Для этой цели Физо применил вращающееся колесо. В более поздних измерениях применялись вращающиеся зеркала и электронные прерыватели. Метод вращающегося зеркала был предложен Уитстоном в 1834 г. и его использовал Фуко в 1860 г. Позднее в течение многих лет Майкельсон систематически занимался усовершенствованием этого метода. Среднее значение c, полученное примерно из 200 измерений Майкельсона, равно 299 796 *км/сек*. Метод оптического прерывателя, использующий

---

*) Понятие скорости электромагнитной волны имеет определенный смысл лишь в связи с волнами очень простого вида, например плоскими волнами. Очевидно, что v не является скоростью распространения в случае произвольного решения уравнений (7), поскольку эти уравнения допускают также решения в виде стоячих волн.

В настоящем вводном разделе предполагается, что читатель знаком с понятием плоской волны, и мы рассматриваем v как скорость ее движения. Математическое представление плоской волны приведено в §§ 1.3 и 1.4.

**) Однако в п. 4.8.1 будет кратко рассмотрена чувствительность глаза человека к различным цветам.

***) Методы, использовавшиеся для определения скорости света, описаны, например, в [8]. Подробный анализ результатов, полученных различными методами, изложен также в [9].

****) Такие измерения дают, по существу, групповую скорость (см п 1.3.4) Различие между групповой и фазовой скоростями света в воздухе при обычных температуре и давлении примерно равно 1 : 50 000.

ячейку Керра, был развит Каролусом и Миттельштетом (1928 г.), Андерсоном (1937 г.) и Хюттелем (1940 г.). Значения c, полученные из этих измерений, находятся в превосходном согласии с теми значениями, которые найдены непрямыми методами, например путем определения отношения значений электрического заряда, измеренных в электростатических и электромагнитных единицах. Так, Роза и Дорсей (1907 г.) нашли последним методом $c = 299\,784$ *км/сек*. Измерения скорости электромагнитных волн в проволоках, выполненные Мерсье (1923 г.), дали значение c, равное 299 782 *км/сек*. Величина, полученная Бирджем [9] на основании тщательного анализа всех имеющихся данных, равна

$$c = 299\,776 \pm 4 \text{ км/сек}. \tag{9}$$

Хорошее согласие значений c, полученных из столь различных экспериментов (причем в некоторых случаях использовалось излучение, частота которого в сотни тысяч раз отличалась от оптических частот), служит замечательным подтверждением теории Максвелла.

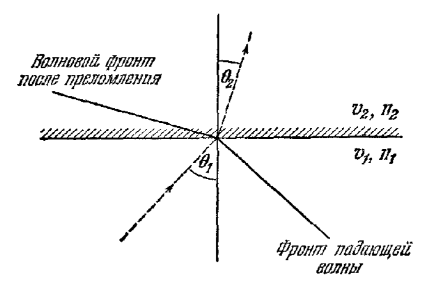


Рис. 1.3. Преломление плоской волны.

Для прозрачных веществ диэлектрическая проницаемость ε обычно больше единицы, а μ практически совпадает с единицей, и следовательно, согласно (8), скорость v меньше скорости в вакууме c. Это было впервые показано экспериментально для случая распространения света в воде в 1850 г. Фуко и Физо.

Как правило, определяется не само значение v, а лишь его отношение к c; для этой цели используется закон преломления. Согласно закону преломления при падении плоской электромагнитной волны на плоскую границу, разделяющую две однородные среды, отношение синуса угла θ_1 между нормалью *) к падающей волне и нормалью к поверхности к синусу угла θ_2 между нормалью к преломленной волне и нормалью к поверхности (рис. 1.3) постоянно и равно отношению скоростей распространения волн в двух средах v_1 и v_2, т. е.

$$\sin\theta_1/\sin\theta_2 = v_1/v_2. \tag{10}$$

Этот результат будет получен в § 1.5. Здесь мы отметим лишь, что он эквивалентен следующему предположению: несмотря на изгиб на границе, волновой фронт остается непрерывным, так что линия пересечения падающей волны с границей движется с такой же скоростью (скажем, v'), как и линия пересечения преломленной волны с границей. Тогда

$$v_1 = v'\sin\theta_1, \qquad v_2 = v'\sin\theta_2. \tag{11}$$

Отсюда, если исключить v', следует (10). Это рассуждение в несколько более развернутой форме часто служит для иллюстрации построения Гюйгенса (см. § 3.3).

Значение постоянного отношения (10) обычно обозначают символом n_{12} и называют *показателем преломления* для преломления из первой среды во вторую. Мы определим также *«абсолютный показатель преломления»* среды n

*) Здесь и ниже термин нормаль к волне означает нормаль к волновому фронту этой волны. Однако для краткости мы будем всюду опускать слова «волновой фронт». (*Прим. ред.*)

как показатель преломления из вакуума в эту среду, т. е.

$$n = c/v. \qquad (12)$$

Если n_1 и n_2— абсолютные показатели преломления двух сред, то относительный показатель преломления n_{12} для преломления из первой среды во вторую равен

$$n_{12} = n_2/n_1 = v_1/v_2. \qquad (13)$$

Из сравнения (12) и (8) получим формулу Максвелла

$$n = \sqrt{\varepsilon\mu} \qquad (14)$$

Для всех веществ, которыми мы будем заниматься, μ практически не отличается от единицы (немагнитные вещества), и поэтому показатель преломления должен равняться квадратному корню из диэлектрической проницаемости, которая, по предположению, является постоянной среды. Вместе с тем хорошо известны эксперименты с призмой, впервые выполненные Ньютоном, которые показывают, что показатель преломления зависит от цвета, т. е. от частоты света. Если мы хотим сохранить формулу Максвелла, то необходимо предположить, что ε не является постоянной величиной, характеризующей среду, а зависит от частоты поля. Зависимость ε от частоты можно объяснить, лишь принимая во внимание атомную структуру вещества. Эта зависимость кратко будет рассмотрена в § 2.3.

Формула Максвелла (с ε — статической диэлектрической проницаемостью) служит хорошим приближенным выражением, например, для газов с простой химической структурой, в которых не происходит существенной дисперсии света, т. е. для веществ с оптическими свойствами, слабо зависящими от цвета света. В табл. 1.1 приведены результаты некоторых ранних измерений

Таблица 1.1

	n (желтый цвет)	$\sqrt{\varepsilon}$
Воздух	1,000294	1,000295
Водород H_2	1,000138	1,000132
Двуокись углерода CO_2	1,000449	1,000473
Окись углерода CO	1,000340	1,000345

Таблица 1.2

	n (желтый цвет)	$\sqrt{\varepsilon}$
Метиловый спирт CH_3OH	1,34	5,7
Этиловый спирт C_2H_5OH	1,36	5,0
Вода H_2O	1,33	9,0

для таких газов, выполненные Л. Больцманом [10]. Формула (14) представляет собой хорошее приближенное выражение и для жидких углеводородов, например, у бензола C_6H_6 $n = 1,482$ для желтого света, а

$$\sqrt{\varepsilon} = 1,489.$$

Вместе с тем для многих твердых тел (например, стекол) и для ряда жидкостей наблюдается сильное отклонение от формулы (14) (табл. 1.2).

§ 1.3. Скалярные волны

В однородной среде в областях, свободных от токов и зарядов, каждая из декартовых компонент $V(\mathbf{r}, t)$ векторов поля удовлетворяет, согласно (1.2.7), однородному волновому уравнению

$$\nabla^2 V - \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} = 0. \qquad (1)$$

Ниже мы кратко исследуем простейшее решение этого уравнения.

1.3.1. Плоские волны. Пусть $\mathbf{r}(x, y, z)$ — радиус-вектор точки P, а $\mathbf{s}(s_x, s_y, s_z)$ — единичный вектор с фиксированным направлением. Говорят, что любое решение уравнения (1) вида

$$V = V(\mathbf{r}\cdot\mathbf{s}, t) \tag{2}$$

представляет *плоскую волну*, так как в каждый момент времени величина V постоянна в плоскостях

$$\mathbf{r}\cdot\mathbf{s} = \text{const},$$

которые перпендикулярны к единичному вектору $\mathbf{s}$.

Удобно выбрать новое положение декартовых осей $O\xi$, $O\eta$, $O\zeta$ так, чтобы ось $O\zeta$ была направлена по $\mathbf{s}$. Тогда (рис. 1.4)

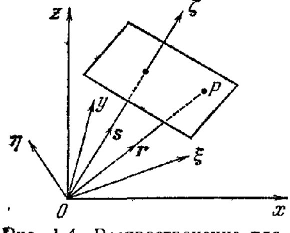


Рис. 1.4. Распространение плоской волны

$$\mathbf{r}\cdot\mathbf{s} = \zeta \tag{3}$$

и

$$\frac{\partial}{\partial x} = s_x \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial y} = s_y \frac{\partial}{\partial \zeta}, \quad \frac{\partial}{\partial z} = s_z \frac{\partial}{\partial \zeta}.$$

Отсюда легко получить

$$\nabla^2 V = \frac{\partial^2 V}{\partial \zeta^2} \tag{4}$$

и, следовательно, волновое уравнение (1) запишется в таком виде

$$\frac{\partial^2 V}{\partial \zeta^2} - \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} = 0. \tag{5}$$

Если мы положим

$$\zeta - vt = p, \qquad \zeta + vt = q, \tag{6}$$

то (5) примет вид

$$\frac{\partial^2 V}{\partial p\, \partial q} = 0. \tag{7}$$

Общим решением этого уравнения служит

$$V = V_1(p) + V_2(q) = V_1(\mathbf{r}\cdot\mathbf{s} - vt) + V_2(\mathbf{r}\cdot\mathbf{s} + vt), \tag{8}$$

где V_1 и V_2 — произвольные функции.

Мы видим, что аргумент функции V_1 не изменяется при замене (ζ, t) на $(\zeta + v\tau, t + \tau)$, где τ произвольно. Следовательно, V_1 представляет возмущение, которое распространяется со скоростью v в положительном направлении оси ζ. Аналогично $V_2(\zeta + vt)$ — это возмущение, распространяющееся со скоростью v в отрицательном направлении оси ζ.

1.3.2. Сферические волны. Теперь рассмотрим решения, представляющие сферические волны, т. е. решения вида

$$V = V(r, t), \tag{9}$$

где $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$. Используя соотношения $\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} = \frac{x}{r}\frac{\partial}{\partial r}$ и т. д., непосредственным расчетом найдем

$$\nabla^2 V = \frac{1}{r}\frac{\partial^2}{\partial r^2}(rV) \tag{10}$$

и, следовательно, волновое уравнение (1) примет вид

$$\frac{\partial^2}{\partial r^2}(rV) - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}(rV) = 0. \tag{11}$$

Это уравнение совпадает с (5), если в последнем заменить ζ на r и V на rV. Следовательно, решение уравнения (11) можно сразу же записать в виде (8), т. е.

$$V = \frac{V_1(r - vt)}{r} + \frac{V_2(r + vt)}{r}, \qquad (12)$$

где V_1 и V_2— по-прежнему произвольные функции. В правой части равенства (12) первый член представляет сферическую волну, расходящуюся от начала координат, второй — сферическую волну, сходящуюся к началу координат, причем скорость распространения обеих волн равна v.

1.3.3. Гармонические волны. Фазовая скорость. В точке $\mathbf{r}_0$ пространства возмущение, вызванное волной, зависит только от времени, т. е.

$$V(\mathbf{r}_0, t) = F(t). \qquad (13)$$

Из приведенных замечаний о цвете очевидно, что особый интерес представляет периодическая функция F. Поэтому рассмотрим случай, когда F имеет вид

$$F(t) = a\cos(\omega t + \delta). \qquad (14)$$

Величина $a (> 0)$ называется *амплитудой*, а аргумент косинуса $(\omega t + \delta)$ — *фазой*. Величина

$$\nu = \frac{\omega}{2\pi} = \frac{1}{T} \qquad (15)$$

называется *частотой* и представляет число колебаний в секунду. Величина ω называется *угловой* (или циклической) частотой и дает число колебаний в 2π секунд. При замене t на $t + T$ значение функции F остается неизменным, поэтому T является *периодом* колебаний. Волновые функции (т. е. решения волнового уравнения) в форме (14) называют *гармоническими* относительно времени.

Рассмотрим вначале волновую функцию, которая представляет *гармоническую плоскую волну*, распространяющуюся в направлении, заданном единичным вектором $\mathbf{s}$. Согласно п. 1.3.1 она получается при замене в формуле (14) t на $t - (\mathbf{r}\cdot\mathbf{s})/v$, т. е.

$$V(\mathbf{r}, t) = a\cos\left[\omega\left(t - \frac{\mathbf{r}\cdot\mathbf{s}}{v}\right) + \delta\right]. \qquad (16)$$

Уравнение (16) не изменится, если $\mathbf{r}\cdot\mathbf{s}$ заменить на $\mathbf{r}\cdot\mathbf{s} + \lambda$, где

$$\lambda = v\frac{2\pi}{\omega} = vT. \qquad (17)$$

Длина λ называется *длиной волны*. Полезно также определить *приведенную длину волны* λ_0

$$\lambda_0 = cT = n\lambda. \qquad (18)$$

Эта длина волны соответствует распространяющейся *в вакууме* гармонической волне той же частоты. В спектроскопии пользуются также понятием волнового числа *) $\varkappa$, которое определяется как число длин волн *в вакууме*, приходящееся на единицу длины (*см*)

$$\varkappa = \frac{1}{\lambda_0} = \frac{\nu}{c}. \qquad (19)$$

Удобно также ввести векторы $\mathbf{k}_0$ и $\mathbf{k}$, направления которых совпадают с направлением распространения $\mathbf{s}$, а длины соответственно равны

$$k_0 = 2\pi\varkappa = 2\pi/\lambda_0 = \omega/c \qquad (20)$$

и

$$k = nk_0 = 2\pi/\lambda = n\omega/c = \omega/v. \qquad (21)$$

Вектор $\mathbf{k} = k\mathbf{s}$ называется *волновым вектором* или *вектором распространения* в среде, а $\mathbf{k}_0 = k_0\mathbf{s}$ — соответствующим вектором для вакуума.

*) Мы будем называть $\varkappa$ «спектроскопическим волновым числом», а термин «волновое число» оставим для k_0 или k, определяемых равенствами (20) и (21), как это принято в оптике.

Вместо постоянной δ применяют также понятие *длины пути* l, представляющей собой расстояние, на которое удаляется волновой фронт при увеличении фазы на δ, т. е.

$$l=\frac{v}{\omega}\delta=\frac{\lambda}{2\pi}\delta=\frac{\lambda_0}{2\pi n}\delta. \tag{22}$$

Рассмотрим теперь гармонические волны более сложной формы. В общем случае вещественную гармоническую скалярную волну с частотой ω можно определить как вещественное решение волнового уравнения вида

$$V(\mathbf{r},\ t)=a(\mathbf{r})\cos[\omega t-g(\mathbf{r})], \tag{23}$$

где $a(>0)$ и g — вещественные скалярные функции положения. Поверхности

$$g(\mathbf{r})=\text{const} \tag{24}$$

называют *поверхностями постоянной фазы*, или *волновыми поверхностями*. В отличие от предыдущего случая, поверхности постоянной амплитуды волны (23), вообще говоря, не совпадают с поверхностями постоянной фазы. Говорят, что такая волна *неоднородна*.

Расчеты, связанные с гармоническими волнами, упрощаются, если использовать экспоненциальные функции вместо тригонометрических. Уравнение (23) можно записать в виде

$$V(\mathbf{r},\ t)=\mathrm{Re}\,\{U(\mathbf{r})\exp[-i\omega t]\}, \tag{25}$$

где

$$U(\mathbf{r})=a(\mathbf{r})\exp[ig(\mathbf{r})], \tag{26}$$

а символ Re означает, что берется вещественная часть. Подставляя (26) в волновое уравнение (1), мы найдем, что U должно удовлетворять уравнению

$$\nabla^2U+n^2k_0^2U=0. \tag{27}$$

Величину U называют *комплексной амплитудой* волны. В частности, для плоской волны имеем *)

$$g(\mathbf{r})=\omega\left(\frac{\mathbf{r}\cdot\mathbf{s}}{v}\right)-\delta=k(\mathbf{r}\cdot\mathbf{s})-\delta=\mathbf{k}\cdot\mathbf{r}-\delta. \tag{28}$$

Если операции, производимые над V, линейны, то в выражении (25) можно опустить символ Re и оперировать прямо с комплексной функцией. При этом вещественная часть окончательного выражения будет представлять изучаемую физическую величину. Однако если приходится иметь дело с нелинейными операциями, такими, как возведение в квадрат и т. д. (например, при расчетах плотности электрической или магнитной энергии), то, вообще говоря, необходимо взять действительные части и оперировать только с ними **).

В отличие от плоской гармонической волны, волна более общего вида (25) не периодична в пространстве. Однако фаза $\omega t-g(\mathbf{r})$ одинакова для $(\mathbf{r},\ t)$ и $(\mathbf{r}+d\mathbf{r},\ t+dt)$ при условии, что

$$\omega\,dt-(\mathrm{grad}\,g)\cdot d\mathbf{r}=0. \tag{29}$$

Обозначив через $\mathbf{q}$ единичный вектор в направлении $d\mathbf{r}$ и написав $d\mathbf{r}=\mathbf{q}\,ds$, найдем отсюда

$$\frac{ds}{dt}=\frac{\omega}{\mathbf{q}\cdot\mathrm{grad}\,g}. \tag{30}$$

Эта величина минимальна, когда вектор $\mathbf{q}$ перпендикулярен к поверхности по-

*) В случае плоских волн часто выделяют постоянный множитель $\exp(-i\delta)$ и называют комплексной амплитудой только переменную часть $a\exp(i\mathbf{k}\cdot\mathbf{r})$.

**) Это не обязательно, когда нас интересует лишь среднее по времени от квадратичного выражения (см. уравнения (1.4.54) — (1.4.56)).

стоянной фазы, т. е. когда $\mathbf{q} = (\text{grad}\, g)/|\text{grad}\, g|$, причем тогда значение (30) будет равно

$$v^{(p)}(\mathbf{r}) = \frac{\omega}{|\text{grad}\, g|}. \tag{31}$$

Величина $v^{(p)}(\mathbf{r})$ называется *фазовой скоростью* и равна скорости, с которой распространяется каждая поверхность постоянной фазы. Для плоской электромагнитной волны из (28) найдем $\text{grad}\, g = \mathbf{k}$ и, учитывая (21), получим $v^{(p)} = \omega/k = c/\sqrt{\varepsilon\mu}$. Для волн более сложной формы, фазовая скорость $v^{(p)}$ в общем случае отличается от $c/\sqrt{\varepsilon\mu}$ и меняется от точки к точке даже в однородной среде. Однако ниже (см. п. 3.1.2) мы увидим, что при достаточно большой частоте фазовая скорость *приблизительно* равна отношению $c/\sqrt{\varepsilon\mu}$ даже для волн, у которых поверхности постоянной фазы не являются плоскими.

Необходимо отметить, что выражение (30) для ds/dt не является компонентой фазовой скорости в направлении $\mathbf{q}$, т. е. фазовая скорость не ведет себя, как вектор. С другой стороны, величина, обратная ей, т. е. величина

$$\frac{dt}{ds} = \frac{\mathbf{q}\, \text{grad}\, g}{\omega}, \tag{32}$$

как видно из этого выражения, есть компонента вектора $(\text{grad}\, g)/\omega$ в направлении $\mathbf{q}$.

В определенных случаях фазовая скорость может превышать c. Для плоских волн это осуществляется, когда $n = \sqrt{\varepsilon\mu}$ меньше единицы, как в случае диспергирующих сред в областях так называемой аномальной дисперсии *) (см. п. 2.3 4). Согласно теории относительности сигналы не могут распространяться со скоростью, превышающей c. Это означает, что фазовая скорость не может соответствовать скорости распространения сигнала. В самом деле, легко видеть, что фазовую скорость нельзя определить экспериментально, и поэтому следует считать ее лишенной какого-либо прямого физического смысла. Для измерения фазовой скорости необходимо было бы сделать отметку на бесконечной гладкой волне и измерить скорость этой отметки, что, однако, означало бы замену бесконечной гармонической волны другой функцией координат и времени.

1.3.4. Волновые пакеты. Групповая скорость. Монохроматические волны, рассматриваемые в предыдущем разделе,— это идеализация, никогда строго не реализующаяся на практике. Из теоремы Фурье следует, что любую волну $V(\mathbf{r}, t)$ (если она удовлетворяет определенным, очень общим условиям) можно рассматривать как суперпозицию монохроматических волн разных частот, а именно

$$V(\mathbf{r}, t) = \int_0^\infty a_\omega(\mathbf{r}) \cos[\omega t - g_\omega(\mathbf{r})]\, d\omega. \tag{33}$$

Здесь снова удобно воспользоваться комплексным представлением, в котором V отождествляется с вещественной частью соответствующей комплексной волны **)

$$V(\mathbf{r}, t) = \text{Re} \int_0^\infty a_\omega(\mathbf{r}) \exp\{-i[\omega t - g_\omega(\mathbf{r})]\}\, d\omega. \tag{33a}$$

---

*) Проблема распространения электромагнитных сигналов в диспергирующих средах очень подробно исследовалась в работах Зоммерфельда [11] и Бриллюэна [12]. Английский перевод этих статей включен в [13] (см. также [14]).

**) Комплексное представление вещественных полихроматических волн более полно изложено в § 10.2.

Если фурье-амплитуды a_ω заметно отличаются от нуля лишь внутри узкого интервала

$$\bar\omega-\frac{1}{2}\Delta\omega\leqslant\omega\leqslant\bar\omega+\frac{1}{2}\Delta\omega\qquad(\Delta\omega/\bar\omega\ll 1)$$

вблизи средней частоты $\bar\omega$, то волну можно назвать «почти монохроматической». В этом случае мы обычно говорим о *волновой группе*, или *волновом пакете* *).

Для иллюстрации некоторых основных свойств волновой группы рассмотрим вначале волну, которая распространяется вдоль оси z и образована в результате суперпозиции двух плоских монохроматических волн с одинаковыми амплитудами и слегка различными частотами и волновыми числами

$$V(z,t)=a\exp\left[-i(\omega t-kz)\right]+a\exp\left\{-i\left[(\omega+\delta\omega)t-(k+\delta k)z\right]\right\}. \tag{34}$$

В соответствии с принятым ранее соглашением символ Re здесь опущен. Уравнение (34) можно записать в форме

$$V(z,t)=a\left\{\exp\left[\frac{1}{2}i(t\delta\omega-z\delta k)\right]+\exp\left[-\frac{1}{2}i(t\delta\omega-z\delta k)\right]\right\}\times$$
$$\times\exp\left[-i(\bar\omega t-\bar kz)\right]=2a\cos\left[\frac{1}{2}(t\delta\omega-z\delta k)\right]\exp\left[-i(\bar\omega t-\bar kz)\right], \tag{35}$$

где

$$\bar\omega=\omega+\frac{1}{2}\delta\omega,\qquad \bar k=k+\frac{1}{2}\delta k \tag{36}$$

— средняя частота и среднее волновое число соответственно. Можно считать что выражение (35) описывает плоскую волну с частотой $\bar\omega$ и длиной волны $2\pi/\bar k$, распространяющуюся в направлении оси z. Однако амплитуда этой волны не постоянна, а изменяется во времени и пространстве от нуля до значения $2a$

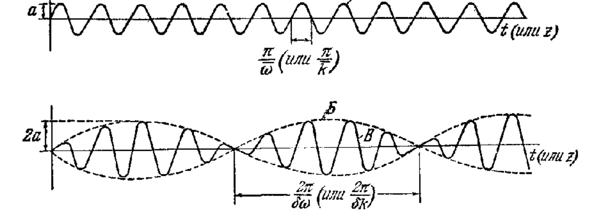


Рис. 1.5. Простая волновая группа.

А — волна $a\cos(\bar\omega t-\bar kz)$; *Б* — волна $2a\cos\left[\frac{1}{2}(t\delta\omega-z\delta k)\right]$; *В* — волновая группа $2a\cos\left[\frac{1}{2}(t\delta\omega-z\delta k)\right]\times\cos(\bar\omega t-\bar kz)$. Абсцисса представляет одну из двух независимых переменных (t или z), тогда как другая сохраняется постоянной

(рис. 1.5), что вызывает хорошо известное явление биений. Соседние максимумы функции, описывающей амплитуду, находятся на расстояниях

$$\left.\begin{aligned}\delta t&=\frac{4\pi}{\delta\omega}\quad\text{при фиксированном } z\\ &\text{или}\\ \delta z&=\frac{4\pi}{\delta k}\quad\text{при фиксированном } t,\end{aligned}\right\} \tag{37}$$

*) Строго говоря, для того чтобы функция V обладала свойствами, которые обычно приписывают волновой группе, необходимо также предположить, что во всем эффективном интервале частот фазовую функцию g_ω можно аппроксимировать линейной функцией ω.

а максимумы функции, связанной с фазой,— на расстояниях

$$\left.\begin{aligned} &\delta t=\frac{2\pi}{\bar\omega} \quad \text{при фиксированном } z \\ \text{или}& \\ &\delta z=\frac{2\pi}{\bar k} \quad \text{при фиксированном } t. \end{aligned}\right\} \qquad (38)$$

Следовательно, поскольку считается, что $\delta\omega/\bar\omega$ и $\delta k/\bar k$ малы по сравнению с единицей, амплитуда будет меняться медленно по сравнению с изменением другого члена

Из (35) вытекает, что плоскости постоянной амплитуды и, в частности, максимумы амплитуды распространяются со скоростью

$$v^{(g)}=\frac{\delta\omega}{\delta k}, \qquad (39)$$

тогда как плоскости постоянной фазы распространяются со скоростью

$$v^{(p)}=\bar\omega/\bar k. \qquad (40)$$

Величина $v^{(g)}$ называется *групповой скоростью* волны. Поскольку V удовлетворяет волновому уравнению, частота ω и волновое число k связаны друг с другом, в среде с показателем преломления n (см. (21))

$$k=n(\omega)\,\omega/c; \qquad (41)$$

здесь показатель преломления n зависит от ω. Уравнение (41) выражает *дисперсию* волны. В недиспергирующей среде n не зависит от ω; в такой среде и фазовая скорость $v^{(p)}$, и групповая скорость $v^{(g)}$ равны c/n. Однако в диспергирующей среде эти две скорости в общем случае различны.

Так как, по предположению, $\delta\omega$ мало, $\delta\omega/\delta k$ можно заменить дифференциальным соотношением $d\omega/dk$; тогда выражение для групповой скорости запишется в виде

$$v^{(g)}=\frac{d\omega}{dk}. \qquad (42)$$

Мы покажем, что фактически это соотношение выполняется при более общих условиях. Рассмотрим одномерную волновую группу

$$V(z,\ t)=\int\limits_{(\Delta\omega)} a_\omega \exp\left[-i(\omega t-kz)\right]d\omega, \qquad (43)$$

где $\Delta\omega$ означает небольшой интервал вблизи средней частоты $\bar\omega$ ($\Delta\omega/\bar\omega \ll 1$), в котором a_ω заметно отличается от нуля. Пусть $\bar k = n(\bar\omega)\,\bar\omega/c$ — соответствующее волновое число. Тогда последнее соотношение можно переписать в форме

$$V(z,\ t)=A(z,\ t)\exp\left[-i(\bar\omega t-\bar kz)\right], \qquad (44)$$

где

$$A(z,\ t)=\int\limits_{(\Delta\omega)} a_\omega \exp\left\{-i\left[(\omega-\bar\omega)t-(k-\bar k)z\right]\right\}d\omega \sim$$

$$\sim \int\limits_{(\Delta\omega)} a_\omega \exp\left\{-i\left[(\omega-\bar\omega)\left\{t-\left(\frac{dk}{d\omega}\right)_{\bar\omega}z\right\}\right]\right\}d\omega, \qquad (45)$$

если $\Delta\omega$ достаточно мало. Снова V можно интерпретировать как плоскую волну с переменной амплитудой, частотой $\bar\omega$ и волновым числом $\bar k$, распространяющуюся в направлении z. Амплитуда $A(z,\ t)$ представляет суперпозицию гармонических волн с частотами $\omega-\bar\omega$. Так как $\Delta\omega/\bar\omega$ мало по сравнению с единицей, A медленно меняется по сравнению с изменением второго члена. В общем случае A комплексно и дает вклад ($\arg A$) в фазу $\bar\omega t - \bar k z$. Как мы

видим, поверхности

$$t=\left(\frac{dk}{d\omega}\right)_{\bar{\omega}} z \tag{46}$$

играют особую роль: на каждой такой поверхности A (z, t) постоянно. Следовательно, скорость перемещения какого-либо значения A, а также максимума $|A|$ определяется, как и раньше, *групповой скоростью*

$$v^{(g)}=\left(\frac{d\omega}{dk}\right)_{\bar{k}}. \tag{47}$$

Легко показать справедливость следующих соотношений между групповой и фазовой скоростями:

$$v^{(g)}=\frac{d}{dk}\left(v^{(p)}k\right)=v^{(p)}+k\frac{dv^{(p)}}{dk}=v^{(p)}-\lambda\frac{dv^{(p)}}{d\lambda}, \tag{48}$$

причем все величины здесь относятся к средней частоте $\bar{\omega}$.

Наконец, рассмотрим трехмерную группу общего вида

$$V(\mathbf{r}, t)=\operatorname{Re}\int\limits_{(\Delta\omega)} a_\omega(\mathbf{r}) \exp\{-i[\omega t-g_\omega(\mathbf{r})]\}\, d\omega. \tag{49}$$

По аналогии с (43) выделим член, соответствующий средней частоте $\bar{\omega}$, и для достаточно малых $\Delta\omega$ напишем

$$V(\mathbf{r}, t)=A(\mathbf{r}, t) \exp\{-i[\bar{\omega} t-g_{\bar{\omega}}(\mathbf{r})]\}, \tag{50}$$

где

$$A(\mathbf{r}, t)=\int\limits_{(\Delta\omega)} a_\omega(\mathbf{r}) \exp\left(-i\{(\omega-\bar{\omega})t-[g_\omega(\mathbf{r})-g_{\bar{\omega}}(\mathbf{r})]\}\right) d\omega \sim$$
$$\sim \int\limits_{(\Delta\omega)} a_\omega(\mathbf{r}) \exp\left(-i\left\{(\omega-\bar{\omega})\left[t-\left(\frac{\partial g(\mathbf{r})}{\partial\omega}\right)_{\bar{\omega}}\right]\right\}\right) d\omega. \tag{51}$$

Выражение (50) представляет волну с частотой $\bar{\omega}$, амплитуда которой A $(\mathbf{r}, t)$ (обычно комплексная) меняется и в пространстве, и во времени, причем ее изменение также медленно по сравнению с изменением второго члена. Можно ожидать, по аналогии с (46), что поверхность

$$t=\left[\frac{\partial g(\mathbf{r})}{\partial\omega}\right]_{\bar{\omega}} \tag{52}$$

будет играть особую роль. Однако теперь амплитудная функция A не обязательно постоянна на каждой такой поверхности, так как здесь фурье-амплитуды a_ω зависят не только от частоты, но и от положения. Значение поверхности, описываемой (52), станет ясным, если мы будем рассматривать абсолютную величину амплитуды $M=|A|$. Имеем

$$M^2(\mathbf{r}, t)=A(\mathbf{r}, t)\cdot A^*(\mathbf{r}, t)=$$
$$=\int\limits_{(\Delta\omega)}\int\limits_{(\Delta\omega)} a_\omega(\mathbf{r})\, a_{\omega'}(\mathbf{r}) \exp\left\{-i(\omega-\omega')\left[t-\left(\frac{\partial g(\mathbf{r})}{\partial\omega}\right)_{\bar{\omega}}\right]\right\} d\omega\, d\omega'. \tag{53}$$

Очевидно, что мнимая часть двойного интеграла равна нулю, так как величина M^2 вещественна. (Формально в этом легко убедиться, если поменять местами независимые переменные ω и ω' и заметить, что при этом мнимая часть подынтегрального выражения меняет знак.) Следовательно,

$$M^2(\mathbf{r}, t)=\int\limits_{(\Delta\omega)}\int\limits_{(\Delta\omega)} a_\omega(\mathbf{r})\, a_{\omega'}(\mathbf{r}) \cos\left\{(\omega-\omega')\left[t-\left(\frac{\partial g(\mathbf{r})}{\partial\omega}\right)_{\bar{\omega}}\right]\right\} d\omega\, d\omega'. \tag{54}$$

Рассматривая какую-нибудь определенную точку $\mathbf{r}=\mathbf{r}_0$ и вспоминая, что a_ω либо положительно, либо равно нулю, мы видим, что M^2 $(\mathbf{r}_0, t)$ достигает максимального значения, когда аргумент косинуса равен нулю, т. е. когда $t=\left(\frac{\partial g(\mathbf{r})}{\partial\omega}\right)_{\bar{\omega}}$. Таким образом, соотношение (52) представляет поверхности,

на которых в момент времени t абсолютное значение амплитуды максимально в указанном выше смысле. Поэтому в общем случае разумно определить групповую скорость трехмерной волновой группы как скорость, с которой перемещаются эти поверхности. Рассмотрим малое смещение $d\mathbf{r} = \mathbf{q}\,ds$, где $\mathbf{q}$ — единичный вектор в направлении нормали к поверхности. Согласно (52) соответствующее изменение δt определяется выражением

$$\delta t = \delta s \left| \operatorname{grad} \left(\frac{\partial g(\mathbf{r})}{\partial \omega} \right)_{\bar{\omega}} \right|, \tag{55}$$

и, следовательно, в общем случае групповая скорость трехмерной группы равна

$$v^{(g)} = \frac{1}{\left| \operatorname{grad} \left(\frac{\partial g}{\partial \omega} \right)_{\bar{\omega}} \right|}. \tag{56}$$

Это выражение нужно сравнивать с выражением для фазовой скорости гармонической волны общего вида (31), т. е. с

$$v^{(p)} = \frac{1}{\left| \operatorname{grad} \frac{g}{\omega} \right|}. \tag{57}$$

В частном случае распространения группы плоских волн в направлении z имеем $g_\omega = kz$, и (56) переходит в (47)

Как ясно из предыдущего, эффективный интервал частот $\Delta\omega$ представляет собой важный параметр, относящийся к волновой группе; по существу эта величина определяет скорость изменения амплитуды и фазы. Если дисперсия среды невелика, волновая группа проходит значительное расстояние без заметного «размытия» При таких обстоятельствах групповая скорость, которую можно считать скоростью распространения группы как целого, является также скоростью распространения энергии (см., например, [15, 16]). Однако в общем случае это неверно. В частности, в области аномальной дисперсии (см. п. 2.3.4) групповая скорость может превысить скорость света или стать отрицательной; в таких случаях она уже не имеет физического смысла.

§ 1.4. Векторные волны

1.4.1. Электромагнитная плоская волна общего вида. Простейшим электромагнитным полем является поле плоской волны. В этом случае, согласно п. 1.3.1, каждая из компонент векторов поля, а следовательно, и векторы $\mathbf{E}$ и $\mathbf{H}$ зависят лишь от переменной $u = \mathbf{r}\cdot\mathbf{s} - vt$, т. е.

$$\mathbf{E} = \mathbf{E}(\mathbf{r}\cdot\mathbf{s} - vt), \qquad \mathbf{H} = \mathbf{H}(\mathbf{r}\cdot\mathbf{s} - vt), \tag{1}$$

здесь $\mathbf{s}$, как и раньше, единичный вектор в направлении распространения.

Обозначим точкой дифференцирование по t и штрихом дифференцирование по переменной u. Тогда

$$\dot{\mathbf{E}} = -v\mathbf{E}', \qquad (\operatorname{rot}\mathbf{E})_x = \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = E_z' s_y - E_y' s_z = (\mathbf{s}\times\mathbf{E}')_x \tag{2}$$

Подставляя эти выражения в уравнения Максвелла (1.1.1) и (1.1.2) с $\mathbf{j} = 0$ и используя материальные уравнения (1.1.10) и (1.1.11), получим

$$\mathbf{s}\times\mathbf{H}' + \frac{\varepsilon v}{c}\mathbf{E}' = 0, \qquad \mathbf{s}\times\mathbf{E}' - \frac{\mu v}{c}\mathbf{H}' = 0. \tag{3}$$

Полагая постоянную интегрирования равной нулю (т. е. пренебрегая постоянным полем во всем пространстве) и, как и раньше, считая $v/c = 1/\sqrt{\varepsilon\mu}$, получим после интегрирования соотношения (3)

$$\mathbf{E} = -\sqrt{\frac{\mu}{\varepsilon}}(\mathbf{s}\times\mathbf{H}), \qquad \mathbf{H} = \sqrt{\frac{\varepsilon}{\mu}}(\mathbf{s}\times\mathbf{E}). \tag{4}$$

Скалярное умножение на s дает

$$\mathbf{E}\cdot\mathbf{s}=\mathbf{H}\cdot\mathbf{s}=0. \tag{5}$$

Это соотношение выражает «поперечность» поля, т. е. оно показывает, что электрический и магнитный векторы лежат в плоскостях, перпендикулярных к направлению распространения. Из соотношений (4) и (5) вытекает, что **E**, **H** и **s** образуют правую ортогональную систему векторов. Из равенств (4) следует также, что

$$\sqrt{\mu}\,H=\sqrt{\varepsilon}\,E, \tag{6}$$

где $E=|\mathbf{E}|$, $H=|\mathbf{H}|$.

Рассмотрим количество энергии, которое протекает в единицу времени через элемент площади, перпендикулярный направлению распространения. Вообразим цилиндр, ось которого параллельна **s**, а площадь поперечного сечения равна единице. Тогда количество энергии, которое протекает через основание цилиндра в единицу времени, равно энергии, содержащейся в части цилиндра длиной v. Следовательно, поток энергии равен vw, где w — плотность энергии. Согласно (6), а также (1.1.31) плотность энергии определяется выражением

$$w=\frac{\varepsilon}{4\pi}E^2=\frac{\mu}{4\pi}H^2. \tag{7}$$

С другой стороны, вектор Пойнтинга, в соответствии с (1.1.38), равен

$$\mathbf{S}=\frac{c}{4\pi}EH\mathbf{s}=\frac{c}{4\pi}\sqrt{\frac{\varepsilon}{\mu}}E^2\mathbf{s}=\frac{c}{4\pi}\sqrt{\frac{\mu}{\varepsilon}}H^2\mathbf{s}. \tag{8}$$

Сравнивая выражения (7) и (8), получим

$$\mathbf{S}=\frac{c}{\sqrt{\varepsilon\mu}}w\mathbf{s}=vw\mathbf{s}. \tag{9}$$

Мы видим, что, в согласии с п. 1.1.4, вектор Пойнтинга представляет собой поток энергии и по величине, и по направлению распространения.

1.4.2. Гармоническая электромагнитная плоская волна. Особый интерес представляет случай плоской волны гармонической во времени, т. е. случай, когда каждая из декартовых компонент векторов **E** и **H** имеет вид

$$a\cos(\tau+\delta)=\mathrm{Re}\{a\exp[-i(\tau+\delta)]\}\qquad (a>0). \tag{10}$$

Здесь τ обозначает переменную часть фазового множителя, т. е.

$$\tau=\omega\left(t-\frac{\mathbf{r}\cdot\mathbf{s}}{v}\right)=\omega t-\mathbf{k}\cdot\mathbf{r}. \tag{11}$$

Выберем ось z в направлении **s**. Тогда отличными от нуля будут лишь x- и y-компоненты **E** и **H**, поскольку, в соответствии с (5), поле поперечно. Ниже мы рассмотрим характер кривой, которую описывает конец электрического вектора в произвольной точке пространства. Эта кривая является геометрическим местом точек, координаты которых (E_x, E_y) равны

$$E_x=a_1\cos(\tau+\delta_1),\quad E_y=a_2\cos(\tau+\delta_2),\quad E_z=0. \tag{12}$$

а. Эллиптическая поляризация. Для того чтобы исключить τ из первых двух уравнений (12), перепишем их в виде

$$E_x/a_1=\cos\tau\cos\delta_1-\sin\tau\sin\delta_1,\qquad E_y/a_2=\cos\tau\cos\delta_2-\sin\tau\sin\delta_2. \tag{13}$$

Следовательно,

$$\left.\begin{aligned}\frac{E_x}{a_1}\sin\delta_2-\frac{E_y}{a_2}\sin\delta_1&=\cos\tau\sin(\delta_2-\delta_1),\\ \frac{E_x}{a_1}\cos\delta_2-\frac{E_y}{a_2}\cos\delta_1&=\sin\tau\sin(\delta_2-\delta_1).\end{aligned}\right\} \tag{14}$$

Возводя в квадрат и складывая, получим

$$\left(\frac{E_x}{a_1}\right)^2+\left(\frac{E_y}{a_2}\right)^2-2\frac{E_x}{a_1}\frac{E_y}{a_2}\cos\delta=\sin^2\delta, \tag{15}$$

где

$$\delta=\delta_2-\delta_1. \tag{16}$$

Соотношение (15) является уравнением конического сечения. Оно имеет форму эллипса, так как соответствующий детерминант неотрицателен, т. е.

$$\begin{vmatrix} \frac{1}{a_1^2} & -\frac{\cos\delta}{a_1a_2} \\ -\frac{\cos\delta}{a_1a_2} & \frac{1}{a_2^2} \end{vmatrix}=\frac{1}{a_1^2a_2^2}(1-\cos^2\delta)=$$

$$=\frac{\sin^2\delta}{a_1^2a_2^2}\geqslant 0.$$

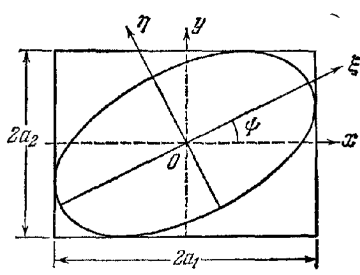


Рис. 1.6. Эллиптически поляризованная волна

Эллипс соответствующий колебанию электрического вектора

Этот эллипс вписан в прямоугольник, стороны которого параллельны осям координат и имеют длины $2a_1$ и $2a_2$ (рис 1.6). Эллипс касается сторон прямоугольника в точках $(\pm a_1, \pm a_2\cos\delta)$ и $(\pm a_1\cos\delta, \pm a_2)$.

В этом случае говорят, что волна, описываемая (12), *эллиптически поляризована*. Легко видеть, что волна, связанная с магнитным вектором, также поляризована эллиптически. Из (4) и (12) следует

$$\left.\begin{aligned} H_x&=-\sqrt{\varepsilon/\mu}\,E_y=-\sqrt{\varepsilon/\mu}\,a_2\cos(\tau+\delta_2),\\ H_y&=\sqrt{\varepsilon/\mu}\,E_x=\sqrt{\varepsilon/\mu}\,a_1\cos(\tau+\delta_1),\\ H_z&=0. \end{aligned}\right\} \tag{17}$$

Конец магнитного вектора описывает эллипс, который вписан в прямоугольник со сторонами длиной $2\sqrt{\varepsilon/\mu}\,a_2$ и $2\sqrt{\varepsilon/\mu}\,a_1$, параллельными осям x и y.

В общем случае оси эллипса не параллельны осям Ox и Oy. Пусть $O\xi$ и $O\eta$ — новые оси, направленные по осям эллипса, а ψ $(0\leqslant\psi<\pi)$ — угол между Ox и направлением главной оси $O\xi$ (см. рис. 1.6). Тогда компоненты E_ξ и E_η будут связаны с E_x и E_y соотношениями

$$E_\xi=E_x\cos\psi+E_y\sin\psi,\qquad E_\eta=-E_x\sin\psi+E_y\cos\psi. \tag{18}$$

Если $2a$ и $2b$ $(a\geqslant b)$ — длины осей эллипса, то уравнение эллипса относительно осей $O\xi$, $O\eta$ будет иметь вид

$$E_\xi=a\cos(\tau+\delta_0),\qquad E_\eta=\pm b\sin(\tau+\delta_0). \tag{19}$$

Наличие двух знаков указывает на возможность двух направлений движения конца электрического вектора, описывающего эллипс.

Чтобы определить a и b, сравним (18) и (19) и используем (13); тогда

$$a(\cos\tau\cos\delta_0-\sin\tau\sin\delta_0)=$$
$$=a_1(\cos\tau\cos\delta_1-\sin\tau\sin\delta_1)\cos\psi+a_2(\cos\tau\cos\delta_2-\sin\tau\sin\delta_2)\sin\psi,$$
$$\pm b(\sin\tau\cos\delta_0+\cos\tau\sin\delta_0)=$$
$$=-a_1(\cos\tau\cos\delta_1-\sin\tau\sin\delta_1)\sin\psi+a_2(\cos\tau\cos\delta_2-\sin\tau\sin\delta_2)\cos\psi.$$

Приравнивая коэффициенты при $\cos\tau$ и $\sin\tau$, получим

$$a\cos\delta_0=a_1\cos\delta_1\cos\psi+a_2\cos\delta_2\sin\psi, \tag{20а}$$
$$a\sin\delta_0=a_1\sin\delta_1\cos\psi+a_2\sin\delta_2\sin\psi, \tag{20б}$$
$$\pm b\cos\delta_0=a_1\sin\delta_1\sin\psi-a_2\sin\delta_2\cos\psi, \tag{21а}$$
$$\pm b\sin\delta_0=-a_1\cos\delta_1\sin\psi+a_2\cos\delta_2\cos\psi. \tag{21б}$$

Возводя в квадрат, складывая (20а) и (20б) и используя (16), находим

$$a^2 = a_1^2 \cos^2\psi + a_2^2 \sin^2\psi + 2a_1a_2 \cos\psi \sin\psi \cos\delta,$$

аналогично из (21а) и (21б) имеем

$$b^2 = a_1^2 \sin^2\psi + a_2^2 \cos^2\psi - 2a_1a_2 \cos\psi \sin\psi \cos\delta. \quad (22)$$

Следовательно,

$$a^2 + b^2 = a_1^2 + a_2^2. \quad (23)$$

Умножим теперь (20а) на (21а), (20б) на (21б) и сложим. Это даст

$$\pm ab = a_1a_2 \sin\delta. \quad (24)$$

Деля (21а) на (20а) и (21б) на (20б), получим

$$\pm\frac{b}{a} = \frac{a_1 \sin\delta_1 \sin\psi - a_2 \sin\delta_2 \cos\psi}{a_1 \cos\delta_1 \cos\psi + a_2 \cos\delta_2 \sin\psi} = \frac{-a_1 \cos\delta_1 \sin\psi + a_2 \cos\delta_2 \cos\psi}{a_1 \sin\delta_1 \cos\psi + a_2 \sin\delta_2 \sin\psi}.$$

Отсюда находим следующее уравнение для ψ:

$$(a_1^2 - a_2^2) \sin 2\psi = 2a_1a_2 \cos\delta \cos 2\psi.$$

Удобно ввести такой вспомогательный угол α $(0 \leqslant \alpha \leqslant \pi/2)$, чтобы

$$a_2/a_1 = \operatorname{tg}\alpha. \quad (25)$$

Тогда предыдущее уравнение примет вид

$$\operatorname{tg} 2\psi = \frac{2a_1a_2}{a_1^2 - a_2^2}\cos\delta = \frac{2\operatorname{tg}\alpha}{1-\operatorname{tg}^2\alpha}\cos\delta,$$

т. е.

$$\operatorname{tg} 2\psi = (\operatorname{tg} 2\alpha)\cos\delta. \quad (26)$$

Из (23) и (24) мы найдем также

$$\pm\frac{2ab}{a^2+b^2} = \frac{2a_1a_2}{a_1^2+a_2^2}\sin\delta = (\sin 2\alpha)\sin\delta. \quad (27)$$

Пусть χ $(-\pi/4 \leqslant \chi \leqslant \pi/4)$ — другой вспомогательный угол, такой что

$$\pm b/a = \operatorname{tg}\chi. \quad (28)$$

Численное значение $\operatorname{tg}\chi$ определяет величину отношения осей эллипса, а знак при χ характеризует два варианта, которые можно использовать при описании эллипса. Перепишем уравнение (27) в виде

$$\sin 2\chi = (\sin 2\alpha)\sin\delta. \quad (29)$$

Полезно кратко просуммировать результаты. Если заданы a_1, a_2 и разность фаз δ, относящиеся к произвольному положению осей, и если α $(0 \leqslant \alpha \leqslant \pi/2)$ — угол, определяемый соотношением

$$\operatorname{tg}\alpha = a_2/a_1, \quad (30)$$

то главные полуоси эллипса a и b и угол ψ $(0 \leqslant \psi < \pi)$, который большая ось образует с осью Ox, находятся из формул

$$a^2 + b^2 = a_1^2 + a_2^2, \quad (31а)$$

$$\operatorname{tg} 2\psi = (\operatorname{tg} 2\alpha)\cos\delta, \quad (31б)$$

$$\sin 2\chi = (\sin 2\alpha)\sin\delta, \quad (31в)$$

где χ $(-\pi/4 < \chi \leqslant \pi/4)$ — вспомогательный угол, определяющий форму и ориентацию эллипса колебаний, а именно:

$$\operatorname{tg}\chi = \pm b/a. \quad (32)$$

Наоборот, если известны длины осей a и b и ориентация эллипса (т. е. заданы a, b и ψ), то эти формулы позволяют найти амплитуды a_1, a_2 и разность фаз δ.

В гл. 14 будут описаны устройства, которые дают возможность определять эти величины прямым способом.

Прежде чем перейти к обсуждению некоторых важных специальных случаев, необходимо сказать несколько слов о терминологии. Мы различаем две поляризации в соответствии с направлением, в котором конец электрического вектора описывает эллипс. По-видимому, естественно было бы называть поляризацию правой или левой в соответствии с тем, образует ли вращение **E** и направление распространения правый или левый винт. Однако принята прямо противоположная терминология: она основана на картине поведения вектора **E**, когда его движение «рассматривается» наблюдателем со стороны положительного направления движения. В настоящей книге мы будем следовать именно такому определению. Итак, будем называть поляризацию *правой*, когда наблюдателю, смотрящему навстречу световому лучу, кажется, что конец электрического вектора описывает эллипс, двигаясь по часовой стрелке. Если для этого случая мы найдем значения величин (12) для двух моментов времени, отличающихся на четверть периода, то увидим, что $\sin\delta > 0$, или, согласно (29), $0 < \chi \leqslant \pi/4$. Для *левой* поляризации справедливо обратное, т. е. наблюдателю, смотрящему навстречу световому лучу, кажется, что электрический вектор описывает эллипс, двигаясь против часовой стрелки. В этом случае $\sin\delta < 0$, так что $-\pi/4 \leqslant \chi < 0$.

По причинам исторического характера, направление магнитного вектора часто называют *направлением поляризации*, а плоскость, в которой лежат магнитный вектор и направление распространения,— *плоскостью поляризации*. Однако такая терминология совсем не общепринята; некоторые авторы определяют эти величины не относительно магнитного, а относительно электрического вектора. Нарушение единообразия возникает частично из-за отсутствия одного-единственного физического понятия, которое можно было бы однозначно считать «световым вектором». Когда особое внимание уделяется физическому действию векторов поля, действительно имеются некоторые основания считать световым вектором вектор **E**. В самом деле, любое действие есть следствие движения элементарных заряженных частиц (электронов, ядер), приведенных в движение электромагнитным полем. В этом случае механическая сила **F**, действующая на частицу со стороны поля, определяется законом Лоренца (см. (1.1.34))

$$\mathbf{F} = e\left(\mathbf{E} + \frac{\mu}{c}[\mathbf{v}\times\mathbf{H}]\right),$$

где e — заряд, **v** — скорость частицы. Следовательно, электрический вектор будет действовать даже на покоящуюся частицу. Вместе с тем магнитный вектор влияет лишь на движущуюся частицу. Однако обычно v/c очень мало по сравнению с единицей, и этим эффектом часто можно пренебречь. Тем не менее «направление поляризации» и «плоскость поляризации», как правило, связывают с магнитным вектором. Причина выбора такой терминологии станет ясной в следующем разделе при рассмотрении поляризации при отражении.

Чтобы избежать путаницы, мы, в соответствии с практикой последнего времени, не будем пользоваться терминами «направление поляризации» или «плоскость поляризации», а будем говорить о *направлении колебаний* и *плоскости колебаний*, чтобы указать направление вектора поля и плоскость, содержащую вектор поля и направление распространения, причем в каждом случае будем оговаривать, о каком именно векторе идет речь.

б. Линейная и круговая поляризации. Наиболее важны два специальных случая, когда эллипс поляризации вырождается либо в прямую, либо в окружность.

Согласно (12) эллипс перейдет в прямую при

$$\delta = \delta_2 - \delta_1 = m\pi \qquad (m = 0, \pm 1, \pm 2, \ldots). \tag{33}$$

Тогда

$$\frac{E_y}{E_x} = (-1)^m \frac{a_2}{a_1}, \tag{34}$$

и мы говорим о *линейной поляризации* *) E. Одну из координатных осей, например x, можно выбрать вдоль этой прямой. Тогда остается лишь одна компонента, а именно E_x. Более того, поскольку электрический и магнитный векторы ортогональны и лежат в плоскости, перпендикулярной к направлению z, компонента H_x тоже исчезает, и, следовательно, вектор **H** линейно поляризован в направлении y.

Другой важный специальный случай — случай *круговой поляризации* волны, когда эллипс вырождается в круг. Ясно, что необходимое условие этого заключается в превращении описанного прямоугольника в квадрат, т. е.

$$a_1 = a_2 = a. \tag{35}$$

Кроме того, одна из компонент вектора **E** должна равняться нулю, когда другая достигает максимального значения; отсюда следует, что

$$\delta = \delta_2 - \delta_1 = \frac{m\pi}{2} \qquad (m = \pm 1,\ \pm 3,\ \pm 5,\ \ldots) \tag{36}$$

и уравнение (15) переходит в уравнение окружности

$$E_x^2 + E_y^2 = a^2. \tag{37}$$

В случае *правой* поляризации $\sin\delta > 0$, так что

$$\delta = \frac{\pi}{2} + 2m\pi \qquad (m = 0,\ \pm 1,\ \pm 2,\ \ldots), \tag{38}$$

$$E_x = a\cos(\tau + \delta_1), \quad E_y = a\cos\left(\tau + \delta_1 + \frac{\pi}{2}\right) = -a\sin(\tau + \delta_1). \tag{39}$$

Для *левой поляризации* $\sin\delta < 0$, так что

$$\delta = -\frac{\pi}{2} + 2m\pi \qquad (m = 0,\ \pm 1,\ \pm 2,\ \ldots), \tag{40}$$

$$E_x = a\cos(\tau + \delta_1), \quad E_y = a\cos\left(\tau + \delta_1 - \frac{\pi}{2}\right) = a\sin(\tau + \delta_1). \tag{41}$$

Если вместо вещественного представления используется комплексное (т. е. если вместо косинусов в (12) написаны экспоненциальные функции), то

$$\frac{E_y}{E_x} = \frac{a_2}{a_1} \exp\left[i(\delta_2 - \delta_1)\right] = \frac{a_2}{a_1} e^{i\delta}, \tag{42}$$

и из значения этого отношения сразу же можно определить характер поляризации.

а) *Линейная поляризация электрической волны* ($\delta = m\pi$, $m = 0$, ± 1, ± 2, ...)

$$\frac{E_y}{E_x} = (-1)^m \frac{a_2}{a_1}.$$

б) *Правая круговая поляризация электрической волны* **) ($a_1 = a_2$, $\delta = \pi/2$)

$$\frac{E_y}{E_x} = \exp(i\pi/2) = i.$$

в) *Левая круговая поляризация электрической волны* ($a_1 = a_2$, $\delta = -\pi/2$)

$$\frac{E_y}{E_x} = \exp(-i\pi/2) = -i.$$

*) Употребляется также менее подходящий термин *плоская поляризация*.

**) Правая по традиционной, а не по естественной терминологии.

В более общем случае можно показать, что для правой эллиптической поляризации мнимая часть отношения E_y/E_x положительна, тогда как для левой эллиптической поляризации она отрицательна.

На рис. 1.7 показаны эллипсы поляризации при разных значениях δ.

в. Характеристика состояния поляризации с помощью параметров Стокса. Для характеристики эллипса поляризации необходимы три независимые величины, например амплитуды a_1 и a_2 и разность фаз δ или малая и большая оси a, b и угол ψ, характеризующий ориентацию эллипса. Для практических целей состояние поляризации удобно охарактеризовать некоторыми параметрами, обладающими одинаковой физической размерностью; они были введены Стоксом в 1852 г. при его исследованиях частично поляризованного света. В общем виде мы определим их позднее (см. п. 10.8 3). Там мы покажем также, что для любой заданной волны эти параметры можно определить из простых экспериментов.

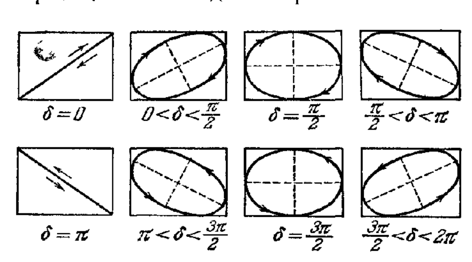


Рис. 1.7. Эллиптическая поляризация при различных значениях разности фаз δ.

Параметрами Стокса для плоской монохроматической волны служат четыре величины:

$$s_0=a_1^2+a_2^2,\qquad s_1=a_1^2-a_2^2,\qquad s_2=2a_1a_2\cos\delta,\qquad s_3=2a_1a_2\sin\delta. \tag{43}$$

Лишь три из них независимы, так как справедливо тождество

$$s_0^2=s_1^2+s_2^2+s_3^2. \tag{44}$$

Очевидно, что параметр s_0 пропорционален интенсивности волны. Параметры s_1, s_2 и s_3 простым образом связаны с углом ψ $(0\leqslant\psi<\pi)$, характеризующим ориентацию эллипса, и углом χ $(-\pi/4\leqslant\chi\leqslant\pi/4)$, характеризующим эллиптичность и направление вращения. Выполняются следующие соотношения:

$$s_1=s_0\cos 2\chi\cos 2\psi, \tag{45а}$$
$$s_2=s_0\cos 2\chi\sin 2\psi, \tag{45б}$$
$$s_3=s_0\sin 2\chi. \tag{45в}$$

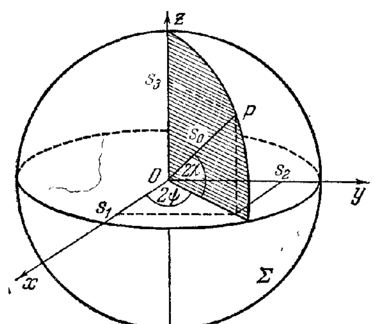


Рис 1 8 Представление состояния поляризации монохроматической волны по Пуанкаре (сфера Пуанкаре).

Выражение (45в) следует из (25) и (29). Для вывода двух других соотношений заметим, что, согласно уравнению, предшествующему (26),

$$s_2=s_1\operatorname{tg}2\psi. \tag{46}$$

Соотношение (45а) получается, если выражения (46) и (45в) подставить в (44). Наконец, (45б) получается при подстановке (45а) в (46).

Выражения (45) указывают простое геометрическое представление различных состояний поляризации: s_1, s_2 и s_3 можно рассматривать как декартовы координаты точки P на сфере Σ радиуса s_0, причем 2χ и 2ψ являются сферическими угловыми координатами этой точки (рис. 1.8). Таким образом, *каждому возможному состоянию поляризации плоской монохроматической волны заданной интенсивности* (s_0 = const) *соответствует одна точка на сфере* Σ *и наоборот.* Так как угол χ положителен или отрицателен в зависимости от того, имеем ли

мы дело с правой или левой поляризацией, то из (45в) следует, что правая поляризация представляется точками на Σ, лежащими выше экваториальной плоскости (плоскости xy), а левая — точками на Σ, лежащими ниже этой плоскости. Далее, для линейно поляризованного света разность фаз равна нулю или целому кратному π; согласно (43) параметр Стокса s_3 равен тогда нулю, так что линейная поляризация представляется точками на экваториальной плоскости. Для круговой поляризации $a_1 = a_2$ и $\delta = \pi/2$ или $\delta = -\pi/2$ в соответствии с тем, имеем ли мы дело с правой или левой поляризацией. Следовательно, правая круговая поляризация представляется северным полюсом ($s_1 = s_2 = 0$, $s_3 = s_0$), а левая поляризация — южным полюсом ($s_1 = s_2 = 0$, $s_3 = -s_0$). Такое геометрическое представление различных состояний поляризации точками на сфере было предложено Пуанкаре [17]. Оно чрезвычайно полезно в кристаллооптике для определения влияния кристаллических сред на состояние поляризации проходящего через них света *). Сфера Σ называется *сферой Пуанкаре*.

1.4.3. Гармонические векторные волны произвольной формы. Основные результаты предыдущих разделов легко распространить на гармонические волны более сложной формы.

Вещественная гармоническая векторная волна общего вида $\mathbf{V}(\mathbf{r}, t)$ является решением векторного волнового уравнения. Проекции $\mathbf{V}$ на оси координат V_x, V_y, V_z представляются выражениями вида (1.3.23), т. е.

$$\left.\begin{aligned} V_x(\mathbf{r}, t) &= a_1(\mathbf{r})\cos[\omega t - g_1(\mathbf{r})], \\ V_y(\mathbf{r}, t) &= a_2(\mathbf{r})\cos[\omega t - g_2(\mathbf{r})], \\ V_z(\mathbf{r}, t) &= a_3(\mathbf{r})\cos[\omega t - g_3(\mathbf{r})], \end{aligned}\right\} \tag{47}$$

где a_s и g_s ($s = 1, 2, 3$) — вещественные функции координат. Для плоской гармонической волны, рассмотренной в предыдущем разделе, величины a постоянны и $g_s(\mathbf{r}) = \mathbf{k}\cdot\mathbf{r} - \delta_s$.

Удобно записать (47) в виде

$$V_x(\mathbf{r}, t) = p_x(\mathbf{r})\cos\omega t + q_x(\mathbf{r})\sin\omega t \quad \text{и т. д.,} \tag{48}$$

где

$$p_x(\mathbf{r}) = a_1(\mathbf{r})\cos g_1(\mathbf{r}), \qquad q_x(\mathbf{r}) = a_1(\mathbf{r})\sin q_1(\mathbf{r}). \tag{49}$$

Если выбрать другое положение осей, то каждая компонента $\mathbf{V}$ в новой системе вновь примет форму (47), так как каждая новая компонента будет линейной комбинацией старых и может поэтому равняться лишь сумме членов с $\cos\omega t$ и $\sin\omega t$.

Мы можем рассматривать (p_x, p_y, p_z) и (q_x, q_y, q_z) как компоненты двух вещественных векторов $\mathbf{p}$ и $\mathbf{q}$. Тогда

$$\mathbf{V}(\mathbf{r}, t) = \mathbf{p}(\mathbf{r})\cos\omega t + \mathbf{q}(\mathbf{r})\sin\omega t. \tag{50}$$

С помощью разложения Фурье можно выразить произвольную векторную волну в виде суперпозиции волн такого типа.

Как и в случае скалярных волн, часто удобно пользоваться комплексным представлением. Запишем (50) в форме

$$\mathbf{V}(\mathbf{r}, t) = \mathrm{Re}\{\mathbf{U}(\mathbf{r})e^{-i\omega t}\}, \tag{51}$$

где $\mathbf{U}$ — комплексный вектор вида

$$\mathbf{U}(\mathbf{r}) = \mathbf{p}(\mathbf{r}) + i\mathbf{q}(\mathbf{r}), \tag{52}$$

а символ Re означает, что берется вещественная часть. Мы можем, как и в соответствующем случае скалярных волн, оперировать прямо с комплексной величиной, опуская символ Re, если операции над $\mathbf{V}$ линейны. Вещественная

*) Примеры, иллюстрирующие этот метод, и ссылки на соответствующую литературу даны в работе [18] (см. также [19, 20]).

часть окончательного выражения тогда представляет рассматриваемую физическую величину.

Действия с комплексными векторами производятся по обычным правилам векторной алгебры и алгебры комплексных чисел. Например, вектор, сопряженный с $\mathbf{U}$, имеет вид

$$\mathbf{U}^* = \mathbf{p} - i\mathbf{q}.$$

Аналогичным образом

$$\mathbf{U}^2 = \mathbf{U}\cdot\mathbf{U} = \mathbf{p}^2 - \mathbf{q}^2 + 2i\mathbf{p}\cdot\mathbf{q}, \qquad \mathbf{U}\cdot\mathbf{U}^* = (\mathbf{p} + i\mathbf{q})\cdot(\mathbf{p} - i\mathbf{q}) = \mathbf{p}^2 + \mathbf{q}^2$$

и т. д.

Для иллюстрации расчетов с комплексными векторами выведем формулы, которые понадобятся нам в дальнейшем, для плотностей энергии и вектора Пойнтинга в случае гармонического электромагнитного поля. Электрический и магнитный векторы запишутся в виде

$$\left.\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \operatorname{Re}\{\mathbf{E}_0(\mathbf{r})\exp(-i\omega t)\} = \\ &= \frac{1}{2}\left[\mathbf{E}_0(\mathbf{r})\exp(-i\omega t) + \mathbf{E}_0^*(\mathbf{r})\exp(i\omega t)\right], \\ \mathbf{H}(\mathbf{r}, t) &= \operatorname{Re}\{\mathbf{H}_0(\mathbf{r})\exp(-i\omega t)\} = \\ &= \frac{1}{2}\left[\mathbf{H}_0(\mathbf{r})\exp(-i\omega t) + \mathbf{H}_0^*(\mathbf{r})\exp(i\omega t)\right], \end{aligned}\right\} \tag{53}$$

где $\mathbf{E}_0$ и $\mathbf{H}_0$ — комплексные векторные функции координат. Так как оптические частоты очень велики (ω порядка 10^{15} *сек*$^{-1}$), нельзя наблюдать мгновенные значения ни одной из таких быстро осциллирующих величин. Можно говорить лишь об их значениях, усредненных по времени за интервал (скажем, $-T' \leqslant t \leqslant T'$), который велик по сравнению с основным периодом $T = 2\pi/\omega$. В частности, средняя по времени плотность электрической энергии равна

$$\langle w_e \rangle = \frac{1}{2T'}\int_{-T'}^{T'} \frac{\varepsilon}{8\pi}\mathbf{E}^2\,dt =$$
$$= \frac{\varepsilon}{8\pi}\frac{1}{2T'}\int_{-T'}^{T'} \frac{1}{4}\left(\mathbf{E}_0^2\exp(-2i\omega t) + 2\mathbf{E}_0\cdot\mathbf{E}_0^* + \mathbf{E}_0^{*2}\exp(2i\omega t)\right)dt.$$

Но

$$\frac{1}{2T'}\int_{-T'}^{T'}\exp(-2i\omega t)\,dt = -\frac{1}{4i\omega T'}\left[\exp(-2i\omega t)\right]_{-T'}^{T'} = \frac{1}{4\pi}\frac{T}{T'}\sin 2\omega T'.$$

Так как, по предположению, интервал T' велик по сравнению с T, величина T/T' мала по сравнению с единицей, и поэтому можно пренебречь интегралом, содержащим $\exp(-2i\omega t)$. Подобным же образом можно пренебречь интегралом, содержащим $\exp(2i\omega t)$, и окончательно мы получим

$$\langle w_e \rangle = \frac{\varepsilon}{16\pi}\mathbf{E}_0\cdot\mathbf{E}_0^*. \tag{54}$$

Аналогично средняя по времени плотность магнитной энергии запишется в виде

$$\langle w_m \rangle = \frac{\mu}{16\pi}\mathbf{H}_0\cdot\mathbf{H}_0^*. \tag{55}$$

Среднее значение вектора Пойнтинга равно

$$\langle \mathbf{S} \rangle = \frac{1}{2T'}\int_{-T'}^{T'}\frac{c}{4\pi}(\mathbf{E}\times\mathbf{H})\,dt = \frac{c}{4\pi}\frac{1}{2T'}\int_{-T'}^{T'}\frac{1}{4}(\mathbf{E}_0\times\mathbf{H}_0\exp(-2i\omega t) +$$
$$+ \mathbf{E}_0\times\mathbf{H}_0^* + \mathbf{E}_0^*\times\mathbf{H}_0 + \mathbf{E}_0^*\times\mathbf{H}_0^*\exp(2i\omega t))\,dt \approx$$
$$\approx \frac{c}{16\pi}(\mathbf{E}_0\times\mathbf{H}_0^* + \mathbf{E}_0^*\times\mathbf{H}_0) = \frac{c}{8\pi}\operatorname{Re}\{\mathbf{E}_0\times\mathbf{H}_0^*\}. \tag{56}$$

Простую форму принимает также закон сохранения энергии. Для непроводящей среды ($\sigma = 0$), где не производится механическая работа, находим, усредняя по времени уравнение (1.1.43),

$$\operatorname{div}\langle \mathbf{S}\rangle = 0. \tag{57}$$

Если проинтегрировать (57) по произвольному объему, который не содержит ни источника, ни поглотителя энергии, и применить теорему Гаусса, то получим

$$\int \langle \mathbf{S}\rangle \cdot \mathbf{n}\, dS = 0. \tag{58}$$

Здесь $\mathbf{n}$ — вектор внешней нормали к граничной поверхности, по которой проводится интегрирование. Таким образом, среднее значение полного потока энергии через любую замкнутую поверхность равно нулю.

Вернемся теперь к гармонической векторной волне общего вида (50) и исследуем поведение $\mathbf{V}$ в точке пространства $\mathbf{r} = \mathbf{r}_0$. В общем случае при изменении времени конец вектора $\mathbf{V}$ описывает эллипс. Поэтому волна вида (50), как и плоская волна, в общем случае также эллиптически поляризована. Чтобы показать это, отметим прежде всего, что с изменением времени конец вектора $\mathbf{V}$ описывает кривую в плоскости, определяемой векторами $\mathbf{p}(\mathbf{r}_0)$ и $\mathbf{q}(\mathbf{r}_0)$. Из периодичности $\mathbf{V}$ следует, что эта кривая должна быть замкнутой. Мы можем положить

$$\mathbf{p} + i\mathbf{q} = (\mathbf{a} + i\mathbf{b})\, e^{i\varepsilon}, \tag{59}$$

где ε — произвольная скалярная величина. Выразим $\mathbf{a}$ и $\mathbf{b}$ через $\mathbf{p}$, $\mathbf{q}$ и ε.

$$\mathbf{a} = \mathbf{p}\cos\varepsilon + \mathbf{q}\sin\varepsilon, \qquad \mathbf{b} = -\mathbf{p}\sin\varepsilon + \mathbf{q}\cos\varepsilon. \tag{60}$$

Выберем ε так, чтобы векторы $\mathbf{a}$ и $\mathbf{b}$ оказались перпендикулярными друг к другу, и допустим, что $|\mathbf{a}| \geqslant |\mathbf{b}|$. Чтобы $\mathbf{a}$ и $\mathbf{b}$ были ортогональны, ε должно удовлетворять уравнению

$$(\mathbf{p}\cos\varepsilon + \mathbf{q}\sin\varepsilon)\cdot(-\mathbf{p}\sin\varepsilon + \mathbf{q}\cos\varepsilon) = 0, \tag{61}$$

т. е.

$$\operatorname{tg} 2\varepsilon = \frac{2\mathbf{p}\cdot\mathbf{q}}{\mathbf{p}^2 - \mathbf{q}^2}. \tag{62}$$

В качестве параметров, определяющих волну, вместо шести компонент векторов $\mathbf{p}$ и $\mathbf{q}$ будем рассматривать пять независимых компонент ортогональных векторов $\mathbf{a}$ и $\mathbf{b}$ и соответствующий фазовый фактор ε. Тогда из соотношений (51), (52) и (59) находим

$$\mathbf{V} = \operatorname{Re}\{(\mathbf{a} + i\mathbf{b})\exp[-i(\omega t - \varepsilon)]\} = \mathbf{a}\cos(\omega t - \varepsilon) + \mathbf{b}\sin(\omega t - \varepsilon). \tag{63}$$

Взяв декартову систему координат с началом в $\mathbf{r}_0$ и с осями x и y, направленными по $\mathbf{a}$ и $\mathbf{b}$, получим

$$V_x = a\cos(\omega t - \varepsilon), \quad V_y = b\sin(\omega t - \varepsilon), \quad V_z = 0. \tag{64}$$

Это уравнение *эллипса* (эллипс *поляризации*)

$$\frac{V_x^2}{a^2} + \frac{V_y^2}{b^2} = 1, \tag{65}$$

длины полуосей которого равны a и b, а направления осей совпадают с осями координат x и y. С помощью элементарной геометрии можно показать, что $\mathbf{p}$ и $\mathbf{q}$ являются парой сопряженных полудиаметров эллипса.

Как и в случае плоских волн, конец вектора может описывать эллипс в двух направлениях, соответствующих левой и правой поляризации; они различаются знаком смешанного произведения $[\mathbf{a}, \mathbf{b}, \nabla\varepsilon] = [\mathbf{p}, \mathbf{q}, \nabla\varepsilon]$.

Длины полуосей эллипса поляризации легко найти из выражений (60) и (62). Согласно (60)

$$a^2 = p^2\cos^2\varepsilon + q^2\sin^2\varepsilon + 2\mathbf{p}\cdot\mathbf{q}\sin\varepsilon\cos\varepsilon =$$
$$= \frac{1}{2}(p^2 + q^2) + \frac{1}{2}(p^2 - q^2)\cos 2\varepsilon + \mathbf{p}\cdot\mathbf{q}\sin 2\varepsilon.$$

Из (62) имеем

$$\sin 2\varepsilon = \frac{2\mathbf{p}\cdot\mathbf{q}}{\sqrt{(p^2-q^2)^2+4(\mathbf{p}\cdot\mathbf{q})^2}}, \qquad \cos 2\varepsilon = \frac{p^2-q^2}{\sqrt{(p^2-q^2)^2+4(\mathbf{p}\cdot\mathbf{q})^2}}.$$

Следовательно,

$$a^2 = \frac{1}{2}\left[p^2+q^2+\sqrt{(p^2-q^2)^2+4(\mathbf{p}\cdot\mathbf{q})^2}\right].$$

Аналогично получим (66)

$$b^2 = \frac{1}{2}\left[p^2+q^2-\sqrt{(p^2-q^2)^2+4(\mathbf{p}\cdot\mathbf{q})^2}\right].$$

Чтобы найти выражение для угла между **a** и **p**, запишем уравнение эллипса в параметрической форме

$$V_x = a\cos\varphi, \qquad V_y = b\sin\varphi, \tag{67}$$

где φ — вспомогательный угол, показанный на рис. 1.9. Из элементарной геометрии известно, что этот угол связан с полярным углом θ точки (V_x, V_y) соотношением

$$\operatorname{tg}\theta = \frac{b}{a}\operatorname{tg}\varphi. \tag{68}$$

Сравнение (64) и (67) показывает, что в данном случае $\varphi = \omega t - \varepsilon$. Согласно (50) $\mathbf{V} = \mathbf{p}$, когда $t = 0$, так что угол φ для вектора **p** равен $-\varepsilon$. Следовательно, угол ψ между **p** и **a** определяется выражением

$$\operatorname{tg}\psi = \frac{b}{a}\operatorname{tg}\varepsilon. \tag{69}$$

Если угол между **p** и **q** обозначить через γ и ввести вспомогательный угол β, определив его соотношением

$$q/p = \operatorname{tg}\beta, \tag{70}$$

то (62) примет вид

$$\operatorname{tg} 2\varepsilon = \frac{2pq}{p^2-q^2}\cos\gamma = \operatorname{tg} 2\beta\cos\gamma. \tag{71}$$

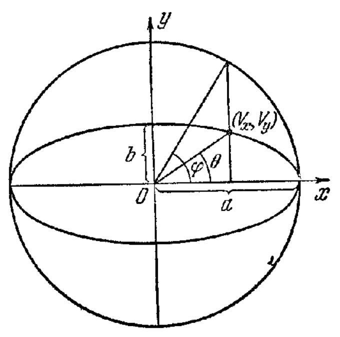


Рис. 1.9. К выводу уравнений (67) и (68).

Суммируем сказанное. Если векторы **p** и **q** заданы, γ — угол между ними, а β — вспомогательный угол, определяемый равенством (70), то главные полуоси эллипса и угол ψ, который большая ось образует с **p**, задаются соотношениями

$$\left.\begin{aligned} a^2 &= \frac{1}{2}\left[p^2+q^2+\sqrt{(p^2-q^2)^2+4p^2q^2\cos^2\gamma}\right],\\ b^2 &= \frac{1}{2}\left[p^2+q^2-\sqrt{(p^2-q^2)^2+4p^2q^2\cos^2\gamma}\right],\\ \operatorname{tg}\psi &= \frac{b}{a}\operatorname{tg}\varepsilon, \end{aligned}\right\} \tag{72}$$

где

$$\operatorname{tg} 2\varepsilon = \operatorname{tg} 2\beta\cos\gamma. \tag{73}$$

Как и для плоских волн, особый интерес представляют два случая, а именно случаи, когда эллипс вырождается в окружность или прямую. Для волны, *поляризованной по кругу*, **a** и **b**, а следовательно, и ε не определены. Согласно (62) для этого необходимо, чтобы

$$\mathbf{p}\cdot\mathbf{q} = \mathbf{p}^2 - \mathbf{q}^2 = 0. \tag{74}$$

Для *линейно поляризованной* волны малая ось равна нулю ($b^2 = 0$), и тогда из

(66) находим

$$p^2q^2 = (\mathbf{p}\cdot\mathbf{q})^2. \tag{75}$$

В заключение мы хотим подчеркнуть, что понятие *«поляризация»* относится к поведению волны в данной точке поля, и поэтому состояние поляризации будет, вообще говоря, неодинаковым в различных точках поля. Таким образом, волна может быть поляризованной линейно или по кругу в одних точках и эллиптически поляризованной в других *). Только в специальных случаях, например для однородной плоской волны, состояние поляризации одинаково во всех точках поля.

§ 1.5. Отражение и преломление плоской волны

В п. 1.1.3 были получены соотношения, которым должны удовлетворять векторы поля на поверхностях, где физические свойства среды претерпевают разрыв. Применим теперь эти формулы к исследованию распространения плоской волны, падающей на плоскую границу, разделяющую две однородные изотропные среды.

1.5.1. Законы отражения и преломления. Если на границу двух однородных сред с разными оптическими свойствами падает плоская волна, она разделяется на две волны: проходящую во вторую среду и отраженную. Существование двух волн вытекает из граничных условий, так как легко видеть, что последние невозможно удовлетворить, если не постулировать наличия как проходящей, так и отраженной волн. Предположим, что эти волны также являются плоскими, и выведем выражения для их амплитуд и направлений распространения.

Плоская волна, распространяющаяся в направлении единичного вектора $\mathbf{s}^{(i)}$ **), полностью определена, если известно поведение возмущения во времени в одной точке пространства, ибо если $\mathbf{F}(t)$ представляет зависимость возмущения от времени в какой-то одной точке, то эта зависимость в другой точке, отстоящей от первой на $\mathbf{r}$, будет $\mathbf{F}[t-(\mathbf{r}\cdot\mathbf{s})/v]$. На границе двух сред вторичные поля будут так же изменяться во времени, как и первичное поле падающей волны. Следовательно, если $\mathbf{s}^{(r)}$ и $\mathbf{s}^{(t)}$ — единичные векторы в направлении распространения отраженной и прошедшей волн, то, приравнивая аргументы трех волновых функций в точке $\mathbf{r}$ на граничной плоскости $z=0$, получим

$$t-\frac{\mathbf{r}\cdot\mathbf{s}^{(i)}}{v_1} = t-\frac{\mathbf{r}\cdot\mathbf{s}^{(r)}}{v_1} = t-\frac{\mathbf{r}\cdot\mathbf{s}^{(t)}}{v_2}, \tag{1}$$

где v_1 и v_2 — скорости распространения в одной и другой средах. Учитывая, что $\mathbf{r}=\{x,\ y,\ 0\}$, находим из (1)

$$\frac{xs_x^{(i)}+ys_y^{(i)}}{v_1} = \frac{xs_x^{(r)}+ys_y^{(r)}}{v_1} = \frac{xs_x^{(t)}+ys_y^{(t)}}{v_2}. \tag{2}$$

Равенства (2) должны выполняться для любых значений x и y на границе, и поэтому

$$\frac{s_x^{(i)}}{v_1} = \frac{s_x^{(r)}}{v_1} = \frac{s_x^{(t)}}{v_2}, \qquad \frac{s_y^{(i)}}{v_1} = \frac{s_y^{(r)}}{v_1} = \frac{s_y^{(t)}}{v_2}. \tag{3}$$

Плоскость, определяемая вектором $\mathbf{s}^{(i)}$ и нормалью к границе, называется *плоскостью падения*. Соотношения (3) показывают, что и $\mathbf{s}^{(r)}$, и $\mathbf{s}^{(t)}$ лежат в этой плоскости.

---

*) Общие свойства гармонических электромагнитных волн произвольной формы, но с линейной поляризацией по крайней мере одного из векторов поля, исследовались в работе [21].

**) Индексы i, r и t относятся соответственно к падающей, отраженной и проходящей (преломленной) волнам.

Считая плоскость xz плоскостью падения и обозначая через θ_i, θ_r и θ_t углы, которые $\mathbf{s}^{(i)}$, $\mathbf{s}^{(r)}$ и $\mathbf{s}^{(t)}$ образуют с осью Oz, получим (рис. 1.10)

$$\left.\begin{array}{lll} s_x^{(i)}=\sin\theta_i, & s_y^{(i)}=0, & s_z^{(i)}=\cos\theta_i, \\ s_x^{(r)}=\sin\theta_r, & s_y^{(r)}=0, & s_z^{(r)}=\cos\theta_r, \\ s_x^{(t)}=\sin\theta_t, & s_y^{(t)}=0, & s_z^{(t)}=\cos\theta_t. \end{array}\right\} \quad (4)$$

Если волна распространяется из первой среды во вторую, компонента вектора $\mathbf{s}$ вдоль оси z положительна; если волна распространяется в противоположном направлении, эта компонента отрицательна, т. е.

$$s_z^{(i)}=\cos\theta_i \geqslant 0, \quad s_z^{(r)}=\cos\theta_r \leqslant 0, \quad s_z^{(t)}=\cos\theta_t \geqslant 0. \quad (5)$$

Подставляя значения (4) в первую систему равенств (3), получим

$$\frac{\sin\theta_i}{v_1}=\frac{\sin\theta_r}{v_1}=\frac{\sin\theta_t}{v_2}. \quad (6)$$

Следовательно, $\sin\theta_r=\sin\theta_i$. Используя (5), мы находим, что $\cos\theta_r=-\cos\theta_i$, поэтому

$$\theta_r=\pi-\theta_i. \quad (7)$$

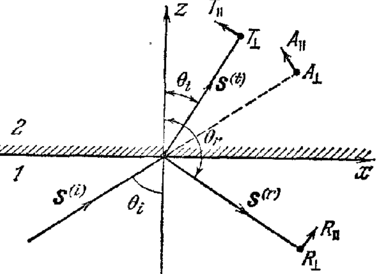


Рис. 1.10. Преломление и отражение плоской волны. Плоскость падения.

Это соотношение вместе с утверждением, что нормаль $\mathbf{s}^{(r)}$ к отраженной волне лежит в плоскости падения, составляет *закон отражения*.

Используя соотношение Максвелла (1.2.14), связывающее показатель преломления и диэлектрическую проницаемость, из (6) найдем также

$$\frac{\sin\theta_i}{\sin\theta_t}=\frac{v_1}{v_2}=\sqrt{\frac{\varepsilon_2\mu_2}{\varepsilon_1\mu_1}}=\frac{n_2}{n_1}=n_{12}. \quad (8)$$

Соотношение $\sin\theta_i/\sin\theta_t=n_2/n_1$ вместе с утверждением, что нормаль $\mathbf{s}^{(t)}$ к преломленной волне лежит в плоскости падения, составляет *закон преломления* (или закон Снеллиуса).

Если $n_2>n_1$, то $n_{12}>1$, и мы говорим, что оптическая плотность второй среды больше, чем первой. В этом случае, учитывая (8), имеем

$$\sin\theta_t=\frac{1}{n_{12}}\sin\theta_i<\sin\theta_i, \quad (9)$$

так что для каждого угла падения существует вещественный угол преломления θ_t. Однако если вторая среда оптически менее плотна, чем первая (т. е. если $n_{12}<1$), то вещественное значение θ_t мы получим лишь для таких углов падения θ_i, для которых $\sin\theta_i \leqslant n_{12}$. Для больших значений θ_i имеет место так называемое *полное внутреннее отражение*. Оно будет рассмотрено ниже в п. 1.5.4.

1.5.2. Формулы Френеля. Здесь мы рассмотрим амплитуды отраженной и преломленной волн. Предположим, что обе среды (однородные и изотропные) обладают нулевой проводимостью и, следовательно, совершенно прозрачны; их магнитные проницаемости фактически будут отличаться от единицы на пренебрежимо малые величины, и поэтому мы положим $\mu_1=\mu_2=1$.

Пусть A — амплитуда электрического вектора поля падающей волны, будем считать ее комплексной величиной с фазой, равной постоянной части аргумента волновой функции. Переменная ее часть имеет вид

$$\tau_i=\omega\left(t-\frac{\mathbf{r}\cdot\mathbf{s}^{(i)}}{v_1}\right)=\omega\left(t-\frac{x\sin\theta_i+z\cos\theta_i}{v_1}\right). \quad (10)$$

Разложим каждый вектор на компоненты — параллельную (снабдим ее индексом $\|$) и перпендикулярную (индекс $\perp$) плоскости падения. Выбор положительных направлений для параллельных компонент указан на рис. 1.10. Перпендикулярные компоненты располагаются перпендикулярно к плоскости рисунка.

Тогда компоненты электрического вектора поля падающей волны запишутся в виде

$$\left.\begin{aligned} E_x^{(i)} &= -A_\| \cos\theta_i \exp(-i\tau_i), \quad E_y^{(i)} = A_\perp \exp(-i\tau_i), \\ E_z^{(i)} &= A_\| \sin\theta_i \exp(-i\tau_i). \end{aligned}\right\} \tag{11}$$

Компоненты магнитного вектора сразу же получаются из соотношения (1.4.4) (при $\mu = 1$)

$$\mathbf{H} = \sqrt{\varepsilon}\,\mathbf{s} \times \mathbf{E}. \tag{12}$$

Отсюда

$$\left.\begin{aligned} H_x^{(i)} &= -A_\perp \cos\theta_i \sqrt{\varepsilon_1} \exp(-i\tau_i), \quad H_y^{(i)} = -A_\| \sqrt{\varepsilon_1} \exp(-i\tau_i), \\ H_z^{(i)} &= A_\perp \sin\theta_i \sqrt{\varepsilon_1} \exp(-i\tau_i). \end{aligned}\right\} \tag{13}$$

Аналогично, если T и R — комплексные амплитуды прошедшей и отраженной волн, то соответствующие компоненты электрического и магнитного векторов равны следующим величинам.

Поле прошедшей волны

$$\left.\begin{aligned} E_x^{(t)} &= -T_\| \cos\theta_t \exp(-i\tau_t); & H_x^{(t)} &= -T_\perp \cos\theta_t \sqrt{\varepsilon_2} \exp(-i\tau_t); \\ E_y^{(t)} &= T_\perp \exp(-i\tau_t); & H_y^{(t)} &= -T_\| \sqrt{\varepsilon_2} \exp(-i\tau_t); \\ E_z^{(t)} &= T_\| \sin\theta_t \exp(-i\tau_t), & H_z^{(t)} &= T_\perp \sin\theta_t \sqrt{\varepsilon_2} \exp(-i\tau_t), \end{aligned}\right\} \tag{14}$$

где

$$\tau_t = \omega\left(t - \frac{\mathbf{r}\cdot\mathbf{s}^{(t)}}{v_2}\right) = \omega\left(t - \frac{x\sin\theta_t + z\cos\theta_t}{v_2}\right). \tag{15}$$

Поле отраженной волны

$$\left.\begin{aligned} E_x^{(r)} &= -R_\| \cos\theta_r \exp(-i\tau_r); & H_x^{(r)} &= -R_\perp \cos\theta_r \sqrt{\varepsilon_1} \exp(-i\tau_r); \\ E_y^{(r)} &= R_\perp \exp(-i\tau_r); & H_y^{(r)} &= -R_\| \sqrt{\varepsilon_1} \exp(-i\tau_r); \\ E_z^{(r)} &= R_\| \sin\theta_r \exp(-i\tau_r), & H_z^{(r)} &= R_\perp \sin\theta_r \sqrt{\varepsilon_1} \exp(-i\tau_r); \end{aligned}\right\} \tag{16}$$

где

$$\tau_r = \omega\left(t - \frac{\mathbf{r}\cdot\mathbf{s}^{(r)}}{v_1}\right) = \omega\left(t - \frac{x\sin\theta_r + z\cos\theta_r}{v_1}\right). \tag{17}$$

Граничные условия (1.1.23) и (1.1.25) требуют, чтобы на границе тангенциальные составляющие векторов $\mathbf{E}$ и $\mathbf{H}$ были непрерывны. Следовательно, должны выполняться соотношения

$$\left.\begin{aligned} E_x^{(i)} + E_x^{(r)} &= E_x^{(t)}, & E_y^{(i)} + E_y^{(r)} &= E_y^{(t)}, \\ H_x^{(i)} + H_x^{(r)} &= H_x^{(t)}, & H_y^{(i)} + H_y^{(r)} &= H_y^{(t)}, \end{aligned}\right\} \tag{18}$$

при этом условия для нормальных компонент $\mathbf{B}$ и $\mathbf{D}$ (1.1.15) и (1.1.19) будут удовлетворяться автоматически. Подставляя в (18) значения всех компонент и используя тот факт, что $\cos\theta_r = \cos(\pi - \theta_i) = -\cos\theta_i$, получим четыре соотношения

$$\left.\begin{aligned} \cos\theta_i (A_\| - R_\|) &= \cos\theta_t T_\|, & A_\perp + R_\perp &= T_\perp, \\ \sqrt{\varepsilon_1}\cos\theta_i (A_\perp - R_\perp) &= \sqrt{\varepsilon_2}\cos\theta_t T_\perp, & \sqrt{\varepsilon_1}(A_\| + R_\|) &= \sqrt{\varepsilon_2} T_\|. \end{aligned}\right\} \tag{19}$$

Заметим, что уравнения (19) делятся на две группы, одна из которых содержит лишь компоненты, параллельные плоскости падения, а другая — только компоненты, перпендикулярные ей. *Следовательно, волны этих двух типов независимы друг от друга.*

Можно решить уравнения (19) относительно компонент отраженной и прошедшей волн, выразив их через компоненты падающей волны. Вновь используя соотношение Максвелла $n=\sqrt{\varepsilon}$, получим

$$T_{\parallel}=\frac{2n_1\cos\theta_i}{n_2\cos\theta_i+n_1\cos\theta_t}A_{\parallel},\qquad T_{\perp}=\frac{2n_1\cos\theta_i}{n_1\cos\theta_i+n_2\cos\theta_t}A_{\perp},\tag{20}$$

$$R_{\parallel}=\frac{n_2\cos\theta_i-n_1\cos\theta_t}{n_2\cos\theta_i+n_1\cos\theta_t}A_{\parallel},\qquad R_{\perp}=\frac{n_1\cos\theta_i-n_2\cos\theta_t}{n_1\cos\theta_i+n_2\cos\theta_t}A_{\perp}\tag{21}$$

Уравнения (20) и (21) называются *формулами Френеля*. Впервые они были выведены Френелем в несколько менее общем виде в 1823 г. на основе его теории, рассматривавшей свет как колебания упругой среды. Эти соотношения пишутся обычно в другой форме, которую можно получить из (20) и (21), используя закон преломления (8), а именно в форме

$$T_{\parallel}=\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i+\theta_t)\cos(\theta_i-\theta_t)}A_{\parallel},\qquad T_{\perp}=\frac{2\sin\theta_t\cos\theta_i}{\sin(\theta_i+\theta_t)}A_{\perp},\tag{20a}$$

$$R_{\parallel}=\frac{\operatorname{tg}(\theta_i-\theta_t)}{\operatorname{tg}(\theta_i+\theta_t)}A_{\parallel},\qquad R_{\perp}=-\frac{\sin(\theta_i-\theta_t)}{\sin(\theta_i+\theta_t)}A_{\perp}.\tag{21a}$$

Так как θ_i и θ_t вещественны (случай полного внутреннего отражения пока исключаем), то тригонометрические функции, стоящие в правой части уравнений (20а) и (21а), также вещественны. Следовательно, фаза каждой компоненты отраженной и прошедшей волн либо равна фазе соответствующей компоненты падающей волны, либо отличается от нее на π. Так как знаки $T_{\parallel}$ и $T_{\perp}$ совпадают со знаками $A_{\parallel}$ и $A_{\perp}$, фаза прошедшей волны равна фазе падающей. В случае же отраженной волны фаза будет зависеть от относительных значений θ_i и θ_t. Если оптическая плотность второй среды больше, чем первой ($\varepsilon_2>\varepsilon_1$), то $\theta_t<\theta_i$; поэтому, согласно (21), знаки $R_{\perp}$ и $A_{\perp}$ различны и фазы отличаются друг от друга на π. При тех же обстоятельствах значение $\operatorname{tg}(\theta_i-\theta_t)$ положительно, но знаменатель $\operatorname{tg}(\theta_i+\theta_t)$ становится отрицательным для $\theta_i+\theta_t>\pi/2$, и в этом случае фазы $R_{\parallel}$ и $A_{\parallel}$ отличаются друг от друга *) на π. Аналогичное рассмотрение можно провести для случая, когда вторая среда оптически менее плотна, чем первая.

Для *нормального падения* $\theta_i=0$ и, следовательно, $\theta_t=0$; тогда соотношения (20) и (21) примут вид

$$T_{\parallel}=\frac{2}{n+1}A_{\parallel},\qquad T_{\perp}=\frac{2}{n+1}A_{\perp},\tag{22}$$

$$R_{\parallel}=\frac{n-1}{n+1}A_{\parallel},\qquad R_{\perp}=-\frac{n-1}{n+1}A_{\perp},\tag{23}$$

где $n=n_2/n_1$.

1.5.3. Отражательная и пропускательная способности; поляризация при отражении и преломлении. Рассмотрим теперь, как энергия поля падающей волны распределяется между двумя вторичными полями.

Интенсивность света (снова считаем $\mu=1$), согласно (1.4.8), равна

$$S=\frac{c}{4\pi}\sqrt{\varepsilon}E^2=\frac{cn}{4\pi}E^2.\tag{24}$$

*) Из (11) и (16) следует, что в плоскости $z=0$

$$\frac{E_y^{(r)}}{E_y^{(i)}}=\frac{R_{\perp}}{A_{\perp}},\qquad \frac{E_x^{(r)}}{E_x^{(i)}}=-\frac{R_{\parallel}}{A_{\parallel}}.$$

Этот результат означает, что в плоскости $z=0$ фазы $E_y^{(r)}$ и $E_y^{(i)}$ отличаются на π, а фазы $E_x^{(r)}$ и $E_x^{(i)}$ равны друг другу. Указанная разница в поведении фаз y- и x-компонент формальна и возникает из-за способа определения угла отражения θ_r (см. рис. 1.10).

Поэтому количество энергии в первичной волне, которое попадает на единицу площади поверхности раздела за 1 *сек*, будет равно

$$J^{(i)} = S^{(i)}\cos\theta_i = \frac{cn_1}{4\pi}|A|^2\cos\theta_i. \tag{25}$$

Для отраженной и преломленной волн энергия, покидающая единицу площади поверхности раздела за 1 *сек*, определяется подобными же выражениями, а именно:

$$J^{(r)} = S^{(r)}\cos\theta_i = \frac{cn_1}{4\pi}|R|^2\cos\theta_i, \qquad J^{(t)} = S^{(t)}\cos\theta_t = \frac{cn_2}{4\pi}|T|^2\cos\theta_t. \tag{26}$$

Отношения

$$\mathcal{R} = \frac{J^{(r)}}{J^{(i)}} = \frac{|R|^2}{|A|^2} \quad \text{и} \quad \mathcal{T} = \frac{J^{(t)}}{J^{(i)}} = \frac{n_2}{n_1}\frac{\cos\theta_t}{\cos\theta_i}\frac{|T|^2}{|A|^2} \tag{27}$$

называют соответственно *отражательной* и *пропускательной способностью* *). Легко проверить, что в соответствии с законом сохранения энергии

$$\mathcal{R} + \mathcal{T} = 1. \tag{28}$$

Отражательная и пропускательная способности зависят от поляризации падающей волны. Их можно выразить через соответствующие отражательную и пропускательную способности для света, поляризованного параллельно и перпендикулярно плоскости падения.

Пусть вектор **E** падающей волны образует с плоскостью падения угол α_i. Тогда

$$A_{\parallel} = A\cos\alpha_i, \qquad A_{\perp} = A\sin\alpha_i. \tag{29}$$

Пусть, далее,

$$J_{\parallel}^{(i)} = \frac{cn_1}{4\pi}|A_{\parallel}|^2\cos\theta_i = J^{(i)}\cos^2\alpha_i, \qquad J_{\perp}^{(i)} = \frac{cn_1}{4\pi}|A_{\perp}|^2\cos\theta_i = J^{(i)}\sin^2\alpha_i \tag{30}$$

и

$$J_{\parallel}^{(r)} = \frac{cn_1}{4\pi}|R_{\parallel}|^2\cos\theta_i, \qquad J_{\perp}^{(r)} = \frac{cn_1}{4\pi}|R_{\perp}|^2\cos\theta_i. \tag{31}$$

Тогда

$$\mathcal{R} = \frac{J^{(r)}}{J^{(i)}} = \frac{J_{\parallel}^{(r)} + J_{\perp}^{(r)}}{J^{(i)}} = \frac{J_{\parallel}^{(r)}}{J_{\parallel}^{(i)}}\cos^2\alpha_i + \frac{J_{\perp}^{(r)}}{J_{\perp}^{(i)}}\sin^2\alpha_i = \mathcal{R}_{\parallel}\cos^2\alpha_i + \mathcal{R}_{\perp}\sin^2\alpha_i, \tag{32}$$

где

$$\left.\begin{aligned} \mathcal{R}_{\parallel} &= \frac{J_{\parallel}^{(r)}}{J_{\parallel}^{(i)}} = \frac{|R_{\parallel}|^2}{|A_{\parallel}|^2} = \frac{\operatorname{tg}^2(\theta_i - \theta_t)}{\operatorname{tg}^2(\theta_i + \theta_t)}, \\ \mathcal{R}_{\perp} &= \frac{J_{\perp}^{(r)}}{J_{\perp}^{(i)}} = \frac{|R_{\perp}|^2}{|A_{\perp}|^2} = \frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)}. \end{aligned}\right\} \tag{33}$$

Подобным же образом получим

$$\mathcal{T} = \frac{J^{(t)}}{J^{(i)}} = \mathcal{T}_{\parallel}\cos^2\alpha_i + \mathcal{T}_{\perp}\sin^2\alpha_i, \tag{34}$$

где

$$\left.\begin{aligned} \mathcal{T}_{\parallel} &= \frac{J_{\parallel}^{(t)}}{J_{\parallel}^{(i)}} = \frac{n_2}{n_1}\frac{\cos\theta_t}{\cos\theta_i}\frac{|T_{\parallel}|^2}{|A_{\parallel}|^2} = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)\cos^2(\theta_i - \theta_t)}, \\ \mathcal{T}_{\perp} &= \frac{J_{\perp}^{(t)}}{J_{\perp}^{(i)}} = \frac{n_2}{n_1}\frac{\cos\theta_t}{\cos\theta_i}\frac{|T_{\perp}|^2}{|A_{\perp}|^2} = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)}. \end{aligned}\right\} \tag{35}$$

*) Как видно из (1.4.8), если $\mu \neq 1$, множитель n_2/n_1 в выражении для $\mathcal{T}$ следует заменить величиной

$$\frac{\sqrt{\varepsilon_2/\mu_2}}{\sqrt{\varepsilon_1/\mu_1}}.$$

Снова можно показать, что

$$\mathcal{R}_{\parallel}+\mathcal{T}_{\parallel}=1, \qquad \mathcal{R}_{\perp}+\mathcal{T}_{\perp}=1. \qquad (36)$$

Для *нормального падения* различие между параллельной и перпендикулярной компонентами исчезает, и из (22), (23) и (27) находим

$$\mathcal{R}=\left(\frac{n-1}{n+1}\right)^2, \qquad \mathcal{T}=\frac{4n}{(n+1)^2}. \qquad (37)$$

Отсюда следует, что

$$\lim_{n\to 1}\mathcal{R}=0, \qquad \lim_{n\to 1}\mathcal{T}=1. \qquad (38)$$

Аналогичные результаты получаются также для предельных значений $\mathcal{R}_{\parallel}$, $\mathcal{T}_{\parallel}$ и $\mathcal{R}_{\perp}$, $\mathcal{T}_{\perp}$. Это легко увидеть из (33) и (35), если учесть, что, согласно закону преломления, $\theta_t \to \theta_i$ при $n \to 1$. Следовательно, чем меньше различие в оптической плотности обеих сред, тем меньше энергии уносится отраженной волной.

Знаменатели в (33) и (35) конечны, за исключением случая $\theta_i+\theta_t=\pi/2$. Тогда $\operatorname{tg}(\theta_i+\theta_t)=\infty$ и, следовательно, $\mathcal{R}_{\parallel}=0$. В этом случае (рис. 1.11) отраженный и преломленный лучи перпендикулярны друг другу, а из закона преломления следует (так как теперь $\sin\theta_t=\sin\{\pi/2-\theta_i\}=\cos\theta_i$), что

$$\operatorname{tg}\theta_i=n. \qquad (39)$$

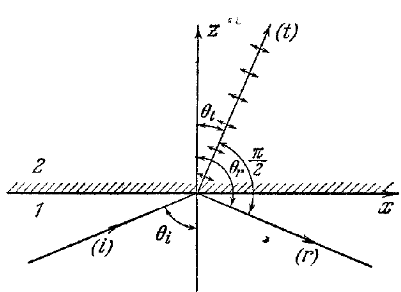


Рис. 1.11. К определению угла полной поляризации (угол Брюстера)

Угол θ_i, определяемый этим выражением, называется *углом полной поляризации* или *углом Брюстера*. Его важность была впервые отмечена в 1815 г. Давидом Брюстером (1781—1868 гг.). *Если свет падает под этим углом, электрический вектор отраженной волны не имеет составляющей в плоскости падения.* Мы обычно говорим в этом случае, что свет поляризован «в плоскости падения». Таким образом, согласно традиционной терминологии, плоскостью поляризации называется плоскость, в которой лежат магнитный вектор и направление распространения. Однако по уже упомянутым в п. 1.4.2 причинам лучше не пользоваться этим термином.

Полученный выше результат, часто называемый *законом Брюстера*, можно пояснить следующим, более прямым рассуждением. Поле падающей волны вызывает колебания электронов в атомах второй среды, которые совершаются в направлении электрического вектора прошедшей волны. Колеблющиеся электроны вызывают отраженную волну, которая распространяется обратно в первую среду. Но линейно колеблющийся электрон излучает в основном в направлении, *перпендикулярном к направлению колебаний* (см. ниже, п. 2.2.3), так что в последнем направлении поток энергии излучения отсутствует. Отсюда следует, что когда отраженный и прошедший лучи перпендикулярны друг другу, то в отраженном луче энергия колебаний в плоскости падения равна нулю.

На рис. 1.12 показана зависимость отражательной способности стекла с показателем преломления 1,52 от угла падения θ_i. Числа над верхней горизонтальной линией относятся к углу преломления θ_t. Нулевое значение $\mathcal{R}_{\parallel}$ на кривой *в* соответствует углу поляризации $\operatorname{arctg} 1{,}52 = 56°40'$.

В оптическом диапазоне показатели преломления по отношению к воздуху обычно порядка 1,5, но в радиодиапазоне они значительно больше, поэтому

там соответственно велики и углы поляризации. Например, для оптических длин волн показатель преломления воды примерно равен 1,3 и угол поляризации 53°. В радиодиапазоне значение показателя преломления достигает примерно 9, а угол поляризации близок к 84°.

Легко видеть, что, согласно (32), кривая *б* на рис. 1.12 соответствует $\alpha = 45°$. Как сейчас будет показано, та же кривая представляет также отражательную способность $\overline{\mathcal{R}}$ для естественного света, т. е. для света, испускаемого нагретым телом. Направление колебаний в естественном свете быстро изменяется беспорядочным, случайным образом. Соответствующую отражательную способность $\overline{\mathcal{R}}$ можно получить путем усреднения по всем направлениям. Так как средние значения $\sin^2\alpha$ и $\cos^2\alpha$ равны 1/2, то для средних значений $J_\parallel^{(i)}$ и $J_\perp^{(i)}$ получим

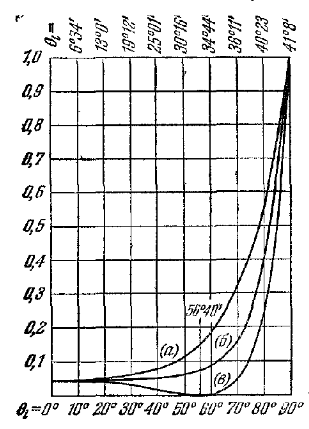


Рис. 1.12. Зависимость отражательной способности от угла падения [22]

а) $\mathcal{R}_\perp$; б) $\frac{1}{2}(\mathcal{R}_\parallel+\mathcal{R}_\perp)$; в) $\mathcal{R}_\parallel$.

$$\overline{J_\parallel^{(i)}} = \overline{J_\perp^{(i)}} = \frac{1}{2} J^{(i)}. \tag{40}$$

Однако для отраженного света обе компоненты в общем случае неодинаковы. В самом деле, используя (40), найдем

$$\left.\begin{aligned} \overline{J_\parallel^{(r)}} &= \frac{1}{2}\frac{\overline{J_\parallel^{(r)}}}{J_\parallel^{(i)}} J^{(i)} = \frac{1}{2}\mathcal{R}_\parallel J^{(i)}, \\ \overline{J_\perp^{(r)}} &= \frac{1}{2}\frac{\overline{J_\perp^{(r)}}}{J_\perp^{(i)}} J^{(i)} = \frac{1}{2}\mathcal{R}_\perp J^{(i)}. \end{aligned}\right\} \tag{41}$$

При этом говорят, что отраженный свет частично поляризован, и *степень его поляризации* P можно определить следующим образом *):

$$P = \left|\frac{\mathcal{R}_\parallel - \mathcal{R}_\perp}{\mathcal{R}_\parallel + \mathcal{R}_\perp}\right|. \tag{42}$$

Отражательная способность определится теперь выражением

$$\overline{\mathcal{R}} = \frac{\overline{J^{(r)}}}{J^{(i)}} = \frac{\overline{J_\parallel^{(r)}} + \overline{J_\perp^{(r)}}}{J^{(i)}} = \frac{1}{2}(\mathcal{R}_\parallel + \mathcal{R}_\perp), \tag{43}$$

и поэтому она по-прежнему будет описываться кривой *б* на рис. 1.12. Степень поляризации теперь можно выразить в виде

$$P = \frac{1}{\overline{\mathcal{R}}}\frac{1}{2}\{|\mathcal{R}_\parallel - \mathcal{R}_\perp|\};$$

выражением в фигурных скобках определяют иногда *поляризованную часть* отраженного света.

Аналогичные результаты можно получить и для проходящего света. Для естественного света мы также найдем

$$\overline{\mathcal{R}} + \overline{\mathcal{T}} = 1. \tag{44}$$

Возвращаясь к случаю линейно поляризованного падающего света, мы видим, что отраженный и прошедший свет останется линейно поляризованным, так как их фазы либо не изменяются, либо изменяются на π. Однако направ-

*) Более общее определение *степени поляризации* дается в п. 10.8.2, где рассматривается также физический смысл этой величины.

ления колебаний в отраженном и проходящем свете изменяются относительно направления колебаний в падающем свете в противоположные стороны. Это можно показать следующим образом.

Угол, который мы обозначили через α, т. е. угол между плоскостью колебаний и плоскостью падения, называют *азимутом* колебания. Мы будем считать его положительным, когда плоскость колебаний поворачивается по часовой стрелке вокруг направления распространения (рис. 1.13). Можно предполагать, что азимут изменяется в пределах от $-\pi/2$ до $\pi/2$. Для падающей, отраженной и прошедшей электрических волн имеем

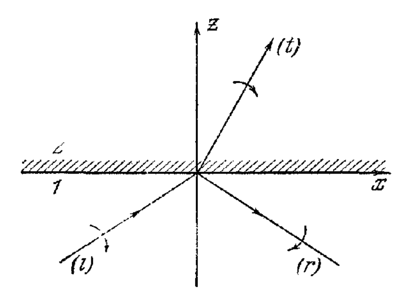


Рис. 1.13. К определению знаков азимутальных углов.

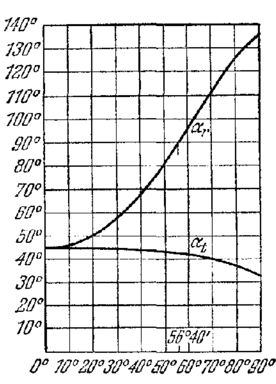


Рис. 1.14. Зависимость азимутальных углов от угла падения [22]

$$\operatorname{tg}\alpha_i = \frac{A_\perp}{A_\parallel}, \qquad \operatorname{tg}\alpha_r = \frac{R_\perp}{R_\parallel}, \qquad \operatorname{tg}\alpha_t = \frac{T_\perp}{T_\parallel}. \tag{45}$$

Используя формулы Френеля (20) и (21), найдем

$$\operatorname{tg}\alpha_r = -\frac{\cos(\theta_i-\theta_t)}{\cos(\theta_i+\theta_t)}\operatorname{tg}\alpha_i, \tag{46}$$

$$\operatorname{tg}\alpha_t = \cos(\theta_i-\theta_t)\operatorname{tg}\alpha_i. \tag{47}$$

Так как $0 \leqslant \theta_i \leqslant \pi/2$, $0 < \theta_t < \pi/2$, то

$$|\operatorname{tg}\alpha_r| \geqslant |\operatorname{tg}\alpha_i|, \tag{48}$$

$$|\operatorname{tg}\alpha_t| \leqslant |\operatorname{tg}\alpha_i|. \tag{49}$$

Знак равенства в соотношении (48) справедлив лишь при нормальном или скользящем падении ($\theta_i = \theta_t = 0$ или $\theta_i = \pi/2$), в соотношении (49) — лишь при нормальном падении. Эти неравенства показывают, что при отражении угол между плоскостью колебаний и плоскостью падения увеличивается, тогда как при преломлении он уменьшается. На рис. 1.14 показано поведение α_r и α_t для $n = 1,52$ и $\alpha_i = 45°$. Мы видим, что когда θ_i равно углу Брюстера 56°40′, то $\alpha_r = 90°$. В самом деле, согласно (46) $\operatorname{tg}\alpha_r = \infty$ (т. е. $\alpha_r = \pi/2$) для $\theta_i + \theta_t = \pi/2$ при любом значении угла α_i.

Из закона Брюстера следует, что свет можно поляризовать, просто заставив его отразиться под углом Брюстера. Один из старейших приборов, основанный на таком принципе,— это так называемый *отражательный прибор* Нюрренберга (Нюрренберг, 1787—1862 гг.). Его основные части — две стеклянные пластинки (рис. 1.15), на которые лучи падают под углом Брюстера. Первая пластинка играет роль *поляризатора*, т. е. приспособления, создающего линей-

но поляризованный свет из неполяризованного света. Вторая служит *анализатором*, т. е. устройством, которое детектирует линейно поляризованный свет. Однако этот прибор обладает рядом недостатков, из них наиболее существенны сравнительно малая доля света, отраженного под углом Брюстера, и довольно сложный путь лучей через прибор. Предпочтительнее использовать устройства, которые поляризуют падающий свет без изменения направления его распространения. Это можно сделать, например, с помощью *стопы тонких плоскопараллельных пластинок*. Если на стопу падает пучок неполяризованного света, то при каждом преломлении он частично поляризуется, и поэтому можно достичь достаточно высокой степени поляризации даже при небольшом числе пластинок. Для отношения интенсивностей двух компонент волны после прохождения через обе поверхности пластинки имеем

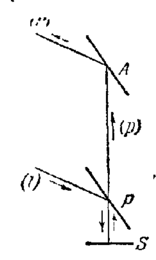


Рис. 1.15. Схема, иллюстрирующая принцип *отражательного прибора Нюрренберга*

P — поляризующая стеклянная пластинка S — отражающее зеркало, A — анализатор, i — падающий пучок, p — поляризованный пучок, r — пучок, отраженный от A

$$\left(\frac{\mathcal{T}_\perp}{\mathcal{T}_\parallel}\right)^2 = \cos^4(\theta_i - \theta_t) < 1, \tag{50}$$

что получается из (35). Таким образом, при выходе из пластинки параллельная компонента преобладает над перпендикулярной, причем степень поляризации тем больше, чем больше угол θ_i. Если θ_i равно углу Брюстера, то $\theta_i + \theta_t = \pi/2$, $\operatorname{tg}\theta_i = n$, и мы получим

$$\left(\frac{\mathcal{T}_\perp}{\mathcal{T}_\parallel}\right)^2 = \sin^4 2\theta_i = \left(\frac{2n}{1+n^2}\right)^4 \tag{51}$$

Для $n = 1,5$ это выражение равно 0,73. Следовательно, если свет проходит, например, через пять пластинок, мы получим отношение, равное $(0,73)^5$, т. е. около 0,2.

Раньше поляризованный свет получали, как правило, с помощью двойного лучепреломления в кристаллах исландского шпата или кварца, как описано в п. 14.4.1. Теперь наиболее удобный метод заключается в использовании так называемых *поляроидных* пленок. Их действие основано на свойстве, известном как *дихроизм*. Вещества, обладающие этим свойством, имеют различные коэффициенты поглощения для света, поляризованного в различных направлениях. Например, можно изготовить пленки из поливинилового спирта с внедренным иодом, которые пропускают почти 80% света, поляризованного в одной плоскости, и менее 1% света, поляризованного в перпендикулярном направлении. Теория этого эффекта кратко будет рассмотрена в п. 14.6.3.

1.5.4. Полное внутреннее отражение. До сих пор мы исключали случай, когда закон преломления

$$\sin\theta_t = \frac{\sin\theta_i}{n_{12}} \tag{52}$$

не дает вещественного значения для угла преломления θ_t. Сейчас мы исследуем этот случай. Он осуществляется при распространении света из оптически более плотной среды в среду с меньшей оптической плотностью, т. е. когда

$$n_{12} = n_2/n_1 = \sqrt{\varepsilon_2\mu_2/\varepsilon_1\mu_1} < 1,$$

при условии, что угол падения θ_i превосходит критическое значение $\bar\theta_i$, определяющееся выражением

$$\sin\bar\theta_i = n_{12} \tag{53}$$

Если $\theta_i = \bar\theta_i$, то $\sin\theta_t = 1$, т. е. $\theta_t = 90°$, так что направление распространения света касательно к поверхности раздела. Если θ_i превышает предельное значе-

ние $\bar{\theta}_i$, свет не входит во вторую среду. Весь падающий свет отражается обратно в первую среду, и мы говорим о *полном внутреннем отражении*.

Тем не менее электромагнитное поле во второй среде не равно нулю, отсутствует лишь поток энергии через границу. Действительно, если в фазовом множителе (15) прошедшей волны мы положим

$$\sin\theta_t = \frac{\sin\theta_i}{n}, \qquad \cos\theta_t = \pm i\sqrt{\frac{\sin^2\theta_i}{n^2} - 1} \tag{54}$$

(индекс 12 у n_{12} опущен), то получим

$$\exp(-i\tau_t) = \exp\left[-i\omega\left(t - \frac{x\sin\theta_i}{nv_2}\right)\right]\exp\left[\mp\frac{\omega z}{v_2}\sqrt{\frac{\sin^2\theta_i}{n^2} - 1}\right]. \tag{55}$$

Выражение (55) описывает неоднородную волну, которая распространяется вдоль поверхности раздела в плоскости падения (т. е. в направлении x) и меняется экспоненциально с изменением расстояния z от этой поверхности. Конечно, физический смысл имеет лишь отрицательный знак перед квадратным корнем в (55), в противном случае при увеличении расстояния амплитуда росла бы неограниченно. Как мы видим, амплитуда очень быстро уменьшается с глубиной проникновения z, причем эффективная глубина проникновения порядка $v_2/\omega = \lambda_2/2\pi$, т. е. порядка длины волны. Эта волна не является поперечной, поскольку, как будет показано ниже, компонента электрического вектора в направлении распространения не равна нулю.

Экспериментальная проверка наличия возмущения во второй (менее плотной) среде представляет довольно трудную задачу, ибо любое устройство, используемое для обнаружения возмущения, будет нарушать граничные условия. Грубое подтверждение можно получить, если поместить вторую преломляющую среду на расстоянии около четверти длины волны от поверхности раздела, на которой происходит полное внутреннее отражение, и наблюдать проникновение излучения во вторую среду. Изящный способ наблюдения этого явления описан в работе [23], где использовались волны длиной 1,25 *см* *).

Чтобы применить формулы Френеля (21а) к случаю полного внутреннего отражения, перепишем их в виде

$$\left.\begin{aligned} R_\parallel &= \frac{\sin\theta_i\cos\theta_i - \sin\theta_t\cos\theta_t}{\sin\theta_i\cos\theta_i + \sin\theta_t\cos\theta_t}A_\parallel, \\ R_\perp &= -\frac{\sin\theta_i\cos\theta_t - \sin\theta_t\cos\theta_i}{\sin\theta_i\cos\theta_t + \sin\theta_t\cos\theta_i}A_\perp. \end{aligned}\right\} \tag{56}$$

Подставив в эти выражения значения величин (54) и помня, что перед квадратным корнем необходимо брать верхний знак, получим

$$\left.\begin{aligned} R_\parallel &= \frac{n^2\cos\theta_i - i\sqrt{\sin^2\theta_i - n^2}}{n^2\cos\theta_i + i\sqrt{\sin^2\theta_i - n^2}}A_\parallel, \\ R_\perp &= \frac{\cos\theta_i - i\sqrt{\sin^2\theta_i - n^2}}{\cos\theta_i + i\sqrt{\sin^2\theta_i - n^2}}A_\perp. \end{aligned}\right\} \tag{57}$$

Следовательно,

$$|R_\parallel| = |A_\parallel|, \qquad |R_\perp| = |A_\perp|, \tag{58}$$

т. е. для каждой компоненты интенсивность света, отраженного при полном внутреннем отражении, равна интенсивности падающего света.

Хотя во второй среде поле отлично от нуля, легко видеть, что поток энергии через поверхность отсутствует. Точнее, можно показать, что хотя в общем

) Проникновение поля во вторую (менее плотную) среду наблюдалось в очень тонких опытах Л. И. Мандельштама еще в 1914 г. [47]. *(Прим. ред.)*

случае компонента вектора Пойнтинга в направлении, нормальном к границе, конечна, ее значение, усредненное по времени, равно нулю. Это означает, что не существует постоянного потока во вторую среду, а энергия течет туда и обратно. Запишем для $z=0$ компоненты поля прошедшей волны вдоль осей x и y и используем (54). (Тут необходимо брать вещественные выражения для **E** и **H**, так как поток энергии является квадратичной функцией компонент.) Отмечая сопряженную комплексную величину звездочкой, получим из (14)

$$E_x^{(t)} = -\frac{1}{2}\left(T_\parallel \cos\theta_t \exp(-i\tau_t^0) + T_\parallel^* \cos^*\theta_t \exp(+i\tau_t^0)\right) =$$

$$= -\frac{i}{2}\sqrt{\frac{\sin^2\theta_i}{n^2}-1}\,\left(T_\parallel \exp(-i\tau_t^0) - T_\parallel^* \exp(+i\tau_t^0)\right),$$

$$E_y^{(t)} = \frac{1}{2}\left(T_\perp \exp(-i\tau_t^0) + T_\perp^* \exp(+i\tau_t^0)\right),$$

$$H_x^{(t)} = -\frac{1}{2}\left(T_\perp \cos\theta_t \sqrt{\varepsilon_2} \exp(-i\tau_t^0) + T_\perp^* \cos^*\theta_t \sqrt{\varepsilon_2} \exp(+i\tau_t^0)\right) =$$

$$= -\frac{i}{2}\sqrt{\varepsilon_2}\sqrt{\frac{\sin^2\theta_i}{n^2}-1}\,\left(T_\perp \exp(-i\tau_t^0) - T_\perp^* \exp(+i\tau_t^0)\right),$$

$$H_y^{(t)} = -\frac{1}{2}\sqrt{\varepsilon_2}\left(T_\parallel \exp(-i\tau_t^0) + T_\parallel^* \exp(+i\tau_t^0)\right),$$

где $\tau_t^0 = \omega\left(t - \frac{x\sin\theta_i}{nv_2}\right)$. Если мы рассмотрим среднее по времени значение величины

$$S_z^{(t)} = \frac{c}{4\pi}\left(E_x^{(t)}H_y^{(t)} - E_y^{(t)}H_x^{(t)}\right)$$

в интервале $-t' \leqslant t \leqslant t'$, где t' велико по сравнению с периодом $T = 2\pi/\omega$, то увидим, что оба члена исчезают при $z=0$. Действительно, один из них содержит множитель

$$\frac{1}{2t'}\int_{-t'}^{t'}\left(T_\parallel^2 \exp(-2i\tau_t^0) - (T_\parallel^*)^2 \exp(+2i\tau_t^0)\right)dt =$$

$$= \left(T_\parallel^2 \exp\left(+\frac{2i\omega x\sin\theta_i}{nv_2}\right) - (T_\parallel^*)^2 \exp\left(-\frac{2i\omega x\sin\theta_i}{nv_2}\right)\right) O\left(\frac{T}{t'}\right),$$

который пренебрежимо мал при $t' \gg T$; другой содержит подобный же множитель с $T_\perp$ вместо $T_\parallel$.

Вместе с тем расчет показывает, что средние по времени значения двух других компонент вектора $\mathbf{S}^{(t)}$ для $z=0$, а именно $S_x^{(t)}$ и $S_y^{(t)}$, в общем случае оказываются конечными. Поэтому энергия не проникает во вторую среду, а течет вдоль поверхности раздела в плоскости падения.

Проведенный выше анализ относился к стационарному состоянию и основывался на предположении, что граничная поверхность и волновые фронты бесконечны. Он не выясняет, как энергия первоначально попала во вторую среду. В реальном эксперименте падающая волна ограничена и в пространстве, и во времени *). В начале процесса небольшое количество энергии будет проникать во вторую среду и создавать там поле.

Наконец, вычислим изменение фаз компонент отраженной и падающей волн. На основании (58) мы можем положить

$$\frac{R_\parallel}{A_\parallel} = e^{i\delta_\parallel}, \qquad \frac{R_\perp}{A_\perp} = e^{i\delta_\perp}. \tag{59}$$

*) Полное внутреннее отражение пучка света конечного поперечного сечения рассматривалось в работе [24].

Согласно (57) каждая из величин $R_{\parallel}/A_{\parallel}$ и $R_{\perp}/A_{\perp}$ имеет форму $z\,(z^*)^{-1}$. Следовательно, если α — аргумент z (т. е. $z = ae^{i\alpha}$, где a и α вещественны), то

$$e^{i\delta} = z\,(z^*)^{-1} = e^{2i\alpha}, \quad \text{т. е.} \quad \operatorname{tg}\frac{\delta}{2} = \operatorname{tg}\alpha,$$

и поэтому

$$\operatorname{tg}\frac{\delta_{\parallel}}{2} = -\frac{\sqrt{\sin^2\theta_i - n^2}}{n^2\cos\theta_i}, \qquad \operatorname{tg}\frac{\delta_{\perp}}{2} = -\frac{\sqrt{\sin^2\theta_i - n^2}}{\cos\theta_i}. \tag{60}$$

Отсюда видно, что обе компоненты испытывают скачки фаз разной величины. Вследствие этого линейно поляризованный свет при полном внутреннем отражении окажется поляризованным эллиптически.

Мы можем теперь записать выражение для относительной разности фаз $\delta = \delta_{\perp} - \delta_{\parallel}$ в виде

$$\operatorname{tg}\frac{\delta}{2} = \frac{\operatorname{tg}\frac{\delta_{\perp}}{2} - \operatorname{tg}\frac{\delta_{\parallel}}{2}}{1 + \operatorname{tg}\frac{\delta_{\perp}}{2}\operatorname{tg}\frac{\delta_{\parallel}}{2}} = \frac{\left(\frac{1}{n^2} - 1\right)\frac{\sqrt{\sin^2\theta_i - n^2}}{\cos\theta_i}}{1 + \frac{\sin^2\theta_i - n^2}{n^2\cos^2\theta_i}},$$

т. е.

$$\operatorname{tg}\frac{\delta}{2} = \frac{\cos\theta_i\sqrt{\sin^2\theta_i - n^2}}{\sin^2\theta_i}. \tag{61}$$

Это выражение обращается в нуль при скользящем падении ($\theta_i = \pi/2$) и при падении под критическим углом $\bar\theta_i$ ($\sin\bar\theta_i = n$). Между этими двумя углами лежит угол, соответствующий максимуму относительной разности фаз. Он определяется уравнением

$$\frac{d}{d\theta_i}\left(\operatorname{tg}\frac{\delta}{2}\right) = \frac{2n^2 - (1+n^2)\sin^2\theta_i}{\sin^3\theta_i\sqrt{\sin^2\theta_i - n^2}} = 0.$$

Это уравнение удовлетворяется, когда

$$\sin^2\theta_i = \frac{2n^2}{1+n^2}. \tag{62}$$

Подставляя (62) в (61), мы получим для максимума δ_m относительной разности фаз δ выражение

$$\operatorname{tg}\frac{\delta_m}{2} = \frac{1-n^2}{2n}. \tag{63}$$

Из (61) видно, что при заданном n существуют два значения угла падения θ_i для каждого значения δ.

Изменение фазы при полном внутреннем отражении можно использовать (как было показано еще Френелем) для получения света, поляризованного по кругу, из линейно поляризованного света. Амплитуды компонент падающего света уравнивают ($|A_{\parallel}| = |A_{\perp}|$), выбирая направление поляризации падающей волны под углом 45° к нормали к плоскости падения (т. е. $\alpha_i = 45°$). Тогда, согласно (58), получим $|R_{\parallel}| = |R_{\perp}|$. Затем n и θ_i подбирают таким образом, чтобы относительная разность фаз δ была равна 90°. Для получения такого значения δ при одном отражении необходимо, согласно (63), чтобы

$$\operatorname{tg}\frac{\pi}{4} = 1 < \frac{1-n^2}{2n}, \quad \text{т. е.} \quad n = n_{12} < \sqrt{2} - 1 = 0{,}414.$$

Отсюда следует, что показатель преломления $n_{21} = 1/n$ более плотной среды относительно менее плотной должен равняться по крайней мере 2,41. Это значение довольно велико, хотя непоглощающие вещества, показатель преломления которых приближается или даже превышает его, существуют. Френель использовал два полных внутренних отражения в стекле. Когда $n_{21} = 1{,}51$, то, согласно (62) и (63), максимальная относительная разность фаз $\delta_m =$

$= 45°56'$ получается при угле падения θ_i, равном $51°20'$. Поэтому значение $\delta = 45°$ можно получить с одним из следующих углов падения:

$$\theta_i = 48°37', \quad \theta_i = 54°37'.$$

Разность фаз в 90° возникает в результате двух последовательных полных внутренних отражений при любом из приведенных углов. Для получения такой разности фаз применяется стеклянный блок, известный как *ромб Френеля* (рис. 1.16).

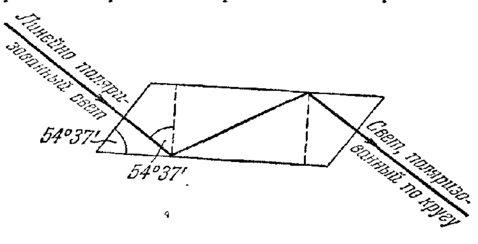


Рис. 1.16. Ромб Френеля.

Разумеется, ромб Френеля можно использовать также для получения эллиптически поляризованного света. В этом случае азимут падающего (линейно поляризованного) света должен отличаться от 45°. Можно также получать посредством ромба Френеля линейно поляризованный свет из света, поляризованного эллиптически.

Измерение критического угла $\bar{\theta}_i$ позволяет удобно и точно определять показатель преломления $n = \sin \bar{\theta}_i$. Приборы, используемые для этой цели, называются *рефрактометрами*.

§ 1.6. Распространение волн в слоистой среде. Теория диэлектрических пленок

Среда, свойства которой постоянны на каждой плоскости, перпендикулярной к фиксированному направлению, называется *слоистой средой*. Если считать это специальное направление осью z декартовой системы координат, то

$$\varepsilon = \varepsilon(z), \qquad \mu = \mu(z). \tag{1}$$

Рассмотрим распространение плоской гармонической электромагнитной волны через такую среду. Это естественное обобщение простого случая, рассмотренного выше.

Теория слоистых сред приобретает важное значение в оптике в связи с *многослойными системами*, т. е. с системами тонких плоскопараллельных пленок. Такие пленки можно изготовлять методом напыления в высоком вакууме, а их толщину можно контролировать с очень большой точностью. Они находят множество полезных приложений. Например, ниже будет показано, что их можно применять в качестве *просветляющих пленок*, т. е. в качестве покрытий, которые уменьшают отражение от данной поверхности. Вместе с тем тонкие пленки при соответствующих условиях будут *увеличивать* отражение. Нанесенные на поверхность стекла пленки можно использовать для разделения пучка; такие устройства применяются в интерферометрии для разделения падающего пучка на две части. При определенных условиях многослойная система может служить фильтром, который пропускает (или отражает) лишь выделенные участки спектра. Многослойные системы употребляются также в качестве поляризаторов.

Вопрос о диэлектрических и металлических пленках очень широко обсуждался в научной литературе. Было предложено много схем для расчета оптических свойств многослойных систем. Мы изложим общую теорию, развитую в превосходных и важных исследованиях Абеле [25] *), и подробно рассмотрим

*) Более детальное изложение вопроса о тонких пленках можно найти в специализированных монографиях, например в [26, 27] или [28].

некоторые специальные случаи, представляющие особый интерес. Естественно, нет необходимости пользоваться общей теорией при рассмотрении проблем, возникающих в случае небольшого числа пленок. Поэтому позднее (см. § 7.6) мы изложим второй, более старый метод, основанный на многократном отражении.

Здесь мы будем заниматься лишь диэлектрической слоистой средой. Распространение этого анализа на проводящие среды проведено в § 13.4.

1.6.1. Основные дифференциальные уравнения. Рассмотрим плоскую гармоническую электромагнитную волну, распространяющуюся через слоистую среду. В частном случае, когда волна поляризована линейно и ее электрический вектор перпендикулярен к плоскости падения, мы будем говорить о *поперечной электрической волне* (обозначаемой TE); если она поляризована линейно и ее магнитный вектор перпендикулярен к плоскости падения, мы будем говорить о *поперечной магнитной волне* (обозначаемой TM) *). Любую произвольно поляризованную плоскую волну можно разложить на две волны, одна из которых является волной TE-типа, а другая — TM-типа. Так как, согласно § 1.5, граничные условия на поверхности раздела для перпендикулярной к ней и параллельной компонент не зависят друг от друга, то эти две волны также будут взаимно независимы. Более того, если поменять местами **E** и **H** и одновременно ε и $-\mu$, то уравнения Максвелла не изменятся. Поэтому любую теорему, относящуюся к TM-волнам, сразу же можно вывести из соответствующего результата для TE-волн с помощью такой замены. Таким образом, достаточно изучить подробно лишь TE-волны.

Возьмем в качестве плоскости падения плоскость **) yz, причем z — направление поперек слоев. Для волны TE-типа $E_y = E_z = 0$, и уравнения Максвелла переходят в следующие шесть скалярных уравнений (зависимость от времени предполагается в виде $\exp(-i\omega t)$):

$$\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}+\frac{i\varepsilon\omega}{c}E_x=0, \quad (1а) \qquad \frac{i\omega\mu}{c}H_x=0, \quad (2а)$$

$$\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}=0, \quad (1б) \qquad \frac{\partial E_x}{\partial z}-\frac{i\omega\mu}{c}H_y=0, \quad (2б)$$

$$\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=0, \quad (1в) \qquad \frac{\partial E_x}{\partial y}+\frac{i\omega\mu}{c}H_z=0. \quad (2в)$$

Эти уравнения показывают, что H_y, H_z и E_x зависят только от y и z. Исключая H_y и H_z из (1а), (2б) и (2в) (или определяя компоненту вдоль оси x волнового уравнения (1.2.5) для **E**), найдем

$$\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial^2 E_x}{\partial z^2}+n^2k_0^2E_x=\frac{d(\ln\mu)}{dz}\frac{\partial E_x}{\partial z}, \quad (3)$$

где

$$n^2=\varepsilon\mu, \qquad k_0=\omega/c=2\pi/\lambda_0. \quad (4)$$

Будем искать решение (3) в виде произведения двух функций, одна из которых зависит лишь от y, а другая только от z:

$$E_x(y, z)=Y(y)\,U(z). \quad (5)$$

Тогда уравнение (3) примет вид

$$\frac{1}{Y}\frac{d^2Y}{dy^2}=-\frac{1}{U}\frac{d^2U}{dz^2}-n^2k_0^2+\frac{d(\ln\mu)}{dz}\frac{1}{U}\frac{dU}{dz}. \quad (6)$$

Левая его часть зависит лишь от y, тогда как правая — лишь от z. Следовательно, (6) может выполняться лишь в том случае, если каждая его часть равна

*) Используются также термины «E-поляризованная» и «H-поляризованная» (см. п. 11.4.1). Нужно упомянуть, что в теории волноводов термины «поперечная электрическая волна» и «поперечная магнитная волна» имеют другое значение.

**) А не плоскость xz, как в предыдущем разделе.

постоянной (скажем, $-K^2$), т. е.

$$\frac{1}{Y}\frac{d^2Y}{dy^2} = -K^2, \tag{7}$$

$$\frac{d^2U}{dz^2} - \frac{d(\ln\mu)}{dz}\frac{dU}{dz} + n^2k_0^2U = K^2U. \tag{8}$$

Удобно положить

$$K^2 = k_0^2\alpha^2. \tag{9}$$

Тогда уравнение (7) дает

$$Y = \text{const}\cdot\exp(ik_0\alpha y),$$

и, следовательно, E_x имеет вид

$$E_x = U(z)\exp[i(k_0\alpha y - \omega t)], \tag{10}$$

где $U(z)$ — функция z (возможно, комплексная). Из (2б) и (2з) мы видим, что выражения для H_y и H_z имеют такую же форму, т. е.

$$H_y = V(z)\exp[i(k_0\alpha y - \omega t)], \tag{11}$$

$$H_z = W(z)\exp[i(k_0\alpha y - \omega t)]. \tag{12}$$

Амплитудные функции U, V и W на основании (1а), (2б) и (2в) связаны следующими уравнениями:

$$V' = ik_0(\alpha W + \varepsilon U), \tag{13а}$$

$$U' = ik_0\mu V, \tag{13б}$$

$$\alpha U + \mu W = 0; \tag{13в}$$

здесь штрих означает дифференцирование по z. Подставляя W из (13в) в (13а), мы получим вместе с (13б) систему из двух дифференциальных уравнений первого порядка *) относительно U и V

$$U' = ik_0\mu V, \quad V' = ik_0\left(\varepsilon - \frac{\alpha^2}{\mu}\right)U. \tag{14}$$

Переходя к уравнениям, содержащим лишь по одной неизвестной функции, окончательно получим следующие линейные дифференциальные уравнения второго порядка для U и V:

$$\frac{d^2U}{dz^2} - \frac{d(\ln\mu)}{dz}\frac{dU}{dz} + k_0^2(n^2 - \alpha^2)U = 0, \tag{15}$$

$$\frac{d^2V}{dz^2} - \frac{d\left[\ln\left(\varepsilon - \frac{\alpha^2}{\mu}\right)\right]}{dz}\frac{dV}{dz} + k_0^2(n^2 - \alpha^2)V = 0. \tag{16}$$

В соответствии с правилом замещения, которое является следствием симметрии уравнений Максвелла, сразу же можно написать, что *для волны ТМ-типа* ($H_y = H_z = 0$) неисчезающие компоненты векторов поля имеют вид

$$H_x = U(z)\exp[i(k_0\alpha y - \omega t)], \tag{17}$$

$$E_y = -V(z)\exp[i(k_0\alpha y - \omega t)], \tag{18}$$

$$E_z = -W(z)\exp[i(k_0\alpha y - \omega t)], \tag{19}$$

причем

$$U' = ik_0\varepsilon V, \quad V' = ik_0\left(\mu - \frac{\alpha^2}{\varepsilon}\right)U, \tag{20}$$

---

*) Форма уравнений (14) совпадает с формой уравнений электрической линии передачи, т. е.

$$dV/dz = -ZI, \qquad dI/dz = -YV,$$

где V — падение напряжения вдоль линии, I — ток в линии, Z — импеданс, а Y — проводимость. Поэтому теорию слоистых сред можно развить совершенно аналогично теории электрических линий передачи, что и было сделано рядом авторов (см., например, [29—30]).

а W и U связаны соотношением

$$\alpha U + \varepsilon W = 0. \tag{21}$$

Функции U и V удовлетворяют следующим линейным дифференциальным уравнениям второго порядка:

$$\frac{d^2U}{dz^2} - \frac{d(\ln \varepsilon)}{dz}\frac{dU}{dz} + k_0^2(n^2-\alpha^2)U = 0, \tag{22}$$

$$\frac{d^2V}{dz^2} - \frac{d\left[\ln\left(\mu - \frac{\alpha^2}{\varepsilon}\right)\right]}{dz}\frac{dV}{dz} + k_0^2(n^2-\alpha^2)V = 0. \tag{23}$$

В общем случае U, V и W являются комплексными функциями z. Поверхности постоянной амплитуды E_x определяются из уравнения

$$|U(z)| = \text{const},$$

поверхности постоянной фазы — из уравнения

$$\varphi(z) + k_0\alpha y = \text{const},$$

где $\varphi(z)$ — фаза U. Оба семейства поверхностей в общем случае не совпадают, так что E_x (и аналогично H_y и H_z) будет неоднородной волной. Для небольшого смещения (dy, dz) вдоль поверхности постоянной фазы имеем

$$\varphi'(z)\,dz + k_0\alpha\,dy = 0;$$

следовательно, если через θ обозначить угол между нормалью к поверхности постоянной фазы и Oz, то получим

$$\operatorname{tg}\theta = -\frac{dz}{dy} = \frac{k_0\alpha}{\varphi'(z)}.$$

В специальном случае однородной плоской волны имеем

$$\varphi(z) = k_0 nz\cos\theta, \qquad \alpha = n\sin\theta. \tag{24}$$

Следовательно, соотношение $\alpha = \text{const}$ с условием (9) можно рассматривать как обобщение *закона преломления Снеллиуса* для слоистых сред.

1.6.2. Характеристическая матрица для слоистой среды. Решения только что выведенных дифференциальных уравнений, подчиняющиеся соответствующим граничным условиям, и различные теоремы, относящиеся к слоистым средам, удобнее представлять в матричной форме. Поэтому перед тем, как рассматривать следствия из наших уравнений, мы кратко изложим основные определения, относящиеся к матрицам.

I. Под матрицей мы понимаем совокупность вещественных или комплексных чисел, расположенных в виде прямоугольной или квадратной таблички

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Символ a_{ij} обозначает элемент в i-й строке j-го столбца. Матрица обозначается символом A или $[a_{ij}]$. Если матрица содержит m строк и n столбцов, то говорят, что это матрица m на n (или $m \times n$-матрица). В специальном случае, когда $m = n$, A называют *квадратной матрицей* порядка m. Если A — квадратная матрица, то определитель, который состоит из таких же элементов в тех же положениях, что и элементы матрицы A, называется *определителем матрицы* A. Он обозначается через $|\mathsf{A}|$ или $|a_{ij}|$. Если $|\mathsf{A}| = 1$, то говорят, что матрица A *унимодулярна*.

По определению две матрицы *равны*, только если они имеют одинаковое число строк m и одинаковое число столбцов n и если равны их соответствующие элементы. Если $\mathsf{A} = [a_{ij}]$ и $\mathsf{B} = [b_{ij}]$ — две матрицы с равным числом строк и равным числом столбцов, то их *сумма* $\mathsf{A}+\mathsf{B}$ определяется как матрица C, элементы которой равны $c_{ij} = a_{ij} + b_{ij}$. Аналогично их *разность* $\mathsf{A}-\mathsf{B}$ определяется как матрица $\mathsf{Д}$ с элементами $d_{ij} = a_{ij} - b_{ij}$.

Матрица, у которой все элементы равны нулю, называется *нулевой матрицей*. Квадратная матрица с элементами $a_{ij} = 0$ при $i \neq j$ и $a_{ii} = 1$ для всех значений i называется *единичной матрицей* и обозначается через I.

Произведение матрицы A и числа λ (вещественного или комплексного) определяется как матрица B с элементами $b_{ij} = \lambda a_{ij}$.

Произведение двух матриц AB определено лишь для случая равенства числа столбцов в A числу строк в B. Если A — это $m \times p$-матрица, а B — это $p \times n$-матрица, то их произведение по определению будет $m \times n$-матрицей с элементами

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

Таким образом, операция умножения двух матриц аналогична умножению определителей равных порядков, когда строка умножается на столбец. В общем случае $\mathsf{AB} \neq \mathsf{BA}$. Например,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

тогда как

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

В специальном случае $\mathsf{AB} = \mathsf{BA}$ матрицы A и B называют *коммутирующими*.

Приведенных определений и свойств матриц достаточно для наших целей, и поэтому мы можем вновь возвратиться к обсуждению распространения электромагнитных волн через слоистую среду.

II. Поскольку функции $U(z)$ и $V(z)$ в п. 1.6.1 удовлетворяют линейным дифференциальным уравнениям второго порядка [типа (15) и (16)], каждую из них можно выразить в виде линейной комбинации двух частных решений, скажем U_1, U_2 и V_1, V_2. Эти частные решения не могут быть произвольными; они должны быть связаны дифференциальными уравнениями первого порядка (14), а именно

$$\left.\begin{aligned} U_1' &= ik_0 \mu V_1, \\ V_1' &= ik_0 \left(\varepsilon - \frac{\alpha^2}{\mu}\right) U_1. \end{aligned}\right\} \qquad \left.\begin{aligned} U_2' &= ik_0 \mu V_2, \\ V_2' &= ik_0 \left(\varepsilon - \frac{\alpha^2}{\mu}\right) U_2. \end{aligned}\right\} \tag{25}$$

Отсюда следует, что

$$V_1 U_2' - U_1' V_2 = 0, \qquad U_1 V_2' - V_1' U_2 = 0,$$

так что

$$\frac{d}{dz}(U_1 V_2 - U_2 V_1) = 0.$$

Это соотношение означает, что *определитель*

$$D = \begin{vmatrix} U_1 & V_1 \\ U_2 & V_2 \end{vmatrix}, \tag{26}$$

соответствующий любым двум произвольным решениям системы (14), постоянен, т. е. что D является инвариантом нашей системы уравнений *).

*) Это следует также из хорошо известного свойства вронскиана для линейных дифференциальных уравнений второго порядка. Более того, нетрудно также показать, что если U_1 известно, то U_2 можно получить с помощью интегрирования из соотношения

$$U_2 = ikDU_1 \int \frac{\mu}{U_1^2}\, dz$$

(см. [25], стр. 603).

Для наших целей наиболее удобно выбрать такое частное решение:

$$U_1 = f(z), \quad U_2 = F(z), \quad V_1 = g(z), \quad V_2 = G(z), \tag{27}$$

чтобы

$$f(0) = G(0) = 0 \quad \text{и} \quad F(0) = g(0) = 1. \tag{28}$$

Тогда решение с

$$U(0) = U_0, \quad V(0) = V_0 \tag{29}$$

можно выразить в виде

$$U = FU_0 + fV_0, \quad V = GU_0 + gV_0,$$

или в матричной форме

$$\mathsf{Q} = \mathsf{N}\mathsf{Q}_0, \tag{30}$$

где

$$\mathsf{Q} = \begin{bmatrix} U(z) \\ V(z) \end{bmatrix}, \quad \mathsf{Q}_0 = \begin{bmatrix} U_0 \\ V_0 \end{bmatrix}, \quad \mathsf{N} = \begin{bmatrix} F(z) & f(z) \\ G(z) & g(z) \end{bmatrix}. \tag{31}$$

Из соотношения $D = \text{const}$ следует, что определитель квадратной матрицы N постоянен. Значение этой постоянной сразу же можно получить, полагая $z = 0$, что дает

$$|\mathsf{N}| = Fg - fG = 1.$$

Обычно удобнее выражать U_0 и V_0 как функции $U(z)$ и $V(z)$. Разрешая относительно U_0 и V_0, получим

$$\mathsf{Q}_0 = \mathsf{M}\mathsf{Q}, \tag{32}$$

где

$$\mathsf{M} = \begin{bmatrix} g(z) & -f(z) \\ -G(z) & F(z) \end{bmatrix}. \tag{33}$$

Эта матрица также унимодулярна, т. е.

$$|\mathsf{M}| = 1. \tag{34}$$

Смысл матрицы M ясен: она связывает x- и y-компоненты электрического (или магнитного) векторов на плоскости $z = 0$ с этими компонентами на произвольной плоскости $z = \text{const}$. Таким образом, мы видим, что для полного определения поля достаточно знать U и V. Следовательно, *для того чтобы узнать, как распространяется плоская монохроматическая волна через слоистую среду, последнюю необходимо охарактеризовать лишь соответствующей унимодулярной* 2×2*-матрицей* M. По этой причине мы будем называть M *характеристической матрицей* слоистой среды. Постоянство определителя $|\mathsf{M}|$ можно показать с помощью закона сохранения энергии *).

Ниже рассматривается форма характеристической матрицы для случаев, представляющих особый интерес.

а. Однородная диэлектрическая пленка. В этом случае величины ε, μ и $n = \sqrt{\varepsilon\mu}$ постоянны. Если θ — угол между нормалью к волне и осью z, то, согласно (24), имеем

$$\alpha = n \sin\theta.$$

Для волны TE-типа получим в соответствии с (15) и (16)

$$\frac{d^2U}{dz^2} + (k_0^2 n^2 \cos^2\theta)\, U = 0, \quad \frac{d^2V}{dz^2} + (k_0^2 n^2 \cos^2\theta)\, V = 0. \tag{35}$$

Легко видеть, что решения этих уравнений, удовлетворяющие соотношениям (14), имеют вид

$$\left.\begin{aligned} U(z) &= A\cos(k_0 nz\cos\theta) + B\sin(k_0 nz\cos\theta), \\ V(z) &= \frac{1}{i}\sqrt{\frac{\varepsilon}{\mu}}\cos\theta\,\{B\cos(k_0 nz\cos\theta) - A\sin(k_0 nz\cos\theta)\}. \end{aligned}\right\} \tag{36}$$

---

*) Чтобы показать это, выразим отражательную и пропускательную способности (51) через матричные элементы. Если, далее, учесть, что для непоглощающей среды характеристическая матрица имеет вид (45), то, как мы видим, закон сохранения $\mathcal{R} + \mathcal{T} = 1$ будет удовлетворяться при условии $|\mathsf{M}| = 1$.

Следовательно, частное решение (27), удовлетворяющее граничным условиям (28), запишется следующим образом:

$$\left.\begin{aligned} U_1 &= f(z) = \frac{i}{\cos\theta}\sqrt{\mu/\varepsilon}\sin(k_0 nz\cos\theta),\\ V_1 &= g(z) = \cos(k_0 nz\cos\theta),\\ U_2 &= F(z) = \cos(k_0 nz\cos\theta),\\ V_2 &= G(z) = i\sqrt{\varepsilon/\mu}\cos\theta\sin(k_0 nz\cos\theta). \end{aligned}\right\} \tag{37}$$

Если мы положим

$$p = \sqrt{\varepsilon/\mu}\cos\theta, \tag{38}$$

то получим характеристическую матрицу в виде

$$M(z) = \begin{bmatrix} \cos(k_0 nz\cos\theta) & -\frac{i}{p}\sin(k_0 nz\cos\theta) \\ -ip\sin(k_0 nz\cos\theta) & \cos(k_0 nz\cos\theta) \end{bmatrix}. \tag{39}$$

Те же уравнения будут справедливы для волны TM-типа, если p заменить на

$$q = \sqrt{\frac{\mu}{\varepsilon}}\cos\theta. \tag{40}$$

б. Слоистая среда, состоящая из тонких однородных пленок. Рассмотрим две смежные слоистые среды, первая из которых занимает пространство от $z = 0$ до $z = z_1$, а вторая — от $z = z_1$ до $z = z_2$. Если $M_1(z)$ и $M_2(z)$ — характеристические матрицы двух сред, то

$$Q_0 = M_1(z)Q(z_1), \quad Q(z_1) = M_2(z_2 - z_1)Q(z_2),$$

так что $Q_0 = M(z_2)Q(z_2)$, где $M(z_2) = M_1(z_1)M_2(z_2 - z_1)$.

Этот результат немедленно можно обобщить на случай непрерывного ряда слоистых сред, расположенных в областях $0 \leqslant z \leqslant z_1$, $z_1 \leqslant z \leqslant z_2$, ..., $z_{N-1} \leqslant z \leqslant z_N$. Если M_1, M_2, ..., M_N — характеристические матрицы сред, то

$$\left.\begin{aligned} Q_0 &= M(z_N)\,Q(z_N),\\ M(z_N) &= M_1(z_1)\,M_2(z_2 - z_1)\cdot\ldots\cdot M_N(z_N - z_{N-1}). \end{aligned}\right\} \tag{41}$$

где

С помощью последнего соотношения легко вывести соответствующее выражение для характеристической матрицы любой слоистой среды *): мы разбиваем эту среду на очень большое число тонких пленок толщиной δz_1, δz_2, δz_3, ..., δz_n. Если их максимальная толщина достаточно мала, можно считать, что ε, μ и n постоянны в каждой пленке. Из (39) видно, что в этом случае характеристическая матрица j-й пленки приближенно равна

$$M_j = \begin{bmatrix} 1 & -\frac{i}{p_j}k_0 n_j \delta z_j\cos\theta_j \\ -ip_j k_0 n_j \delta z_j\cos\theta_j & 1 \end{bmatrix}.$$

Следовательно, характеристическая матрица всей среды, рассматриваемой как совокупность тонких пленок, приблизительно равна

$$M = \prod_{j=1}^{N} M_j = \begin{bmatrix} 1 & -ik_0 B \\ -ik_0 A & 1 \end{bmatrix}, \tag{42}$$

где

$$A = \sum_{j=1}^{N} p_j n_j \delta z_j \cos\theta_j = \sum_{j=1}^{N}\left(\varepsilon_j - \frac{\alpha^2}{\mu_j}\right)\delta z_j,$$

$$B = \sum_{j=1}^{N} \frac{n_j}{p_j}\delta z_j\cos\theta_j = \sum_{j=1}^{N}\mu_j\delta z_j.$$

*) Более подробное рассмотрение слоистых сред с непрерывно меняющимся показателем преломления проведено в работе [31].

Здесь также оставлены лишь члены до первого порядка по δz включительно. Переходя к пределу при $N \to \infty$ таким образом, чтобы максимум $|\delta z_j|$ стремился к нулю, получим

$$\mathsf{M} = \begin{bmatrix} 1 & -ik_0\mathcal{B} \\ -ik_0\mathcal{A} & 1 \end{bmatrix}, \tag{43}$$

где

$$\mathcal{A} = \int \left(\varepsilon - \frac{\alpha^2}{\mu}\right) dz, \quad \mathcal{B} = \int \mu \, dz, \tag{44}$$

а интегрирование проводится по всему интервалу изменения z. Выражение (43) дает первое приближение для характеристической матрицы произвольной слоистой среды. Оставляя в разложении величин $\cos(k_0 n \delta z \cos\theta)$ и $\sin(k_0 n \delta z \cos\theta)$ и в произведении (42) члены более высоких порядков, можно получить следующие приближения *).

Поскольку для непоглощающей среды ε и μ вещественны, то видно также, что *характеристическая матрица непоглощающей слоистой среды имеет вид*

$$\mathsf{M} = \begin{bmatrix} a & ib \\ ic & d \end{bmatrix}, \tag{45}$$

где a, b, c и d вещественны.

1.6.3. Коэффициенты отражения и пропускания. Рассмотрим плоскую волну, падающую на слоистую среду, которая занимает область от $z = 0$ до $z = z_1$ и с обеих сторон граничит с однородными полубесконечными средами. Выведем выражения для амплитуд и интенсивностей отраженной и прошедшей волн **).

Пусть A, R и T обозначают, как и раньше, амплитуды (возможно, комплексные) электрических векторов падающей, отраженной и преломленной волн. Далее пусть ε_1, μ_1, и ε_l, μ_l — диэлектрические и магнитные проницаемости первой и последней сред, а θ_1 и θ_l — углы между нормалями к падающей и прошедшей волнам и направлением оси z (направлением стратификации).

Граничные условия, приведенные в § 1.1, требуют, чтобы тангенциальные компоненты векторов **E** и **H** были непрерывны на каждой из двух поверхностей раздела слоистой среды. Вместе с соотношением (1.4.4)

$$\mathbf{H} = \sqrt{\varepsilon/\mu}\, \mathbf{s} \times \mathbf{E}$$

это приводит к следующим соотношениям для волны TE-типа:

$$\left.\begin{aligned} U_0 &= A + R, & U(z_1) &= T, \\ V_0 &= p_1(A - R), & V(z_1) &= p_l T, \end{aligned}\right\} \tag{46}$$

где

$$p_1 = \sqrt{\varepsilon_1/\mu_1}\cos\theta_1, \quad p_l = \sqrt{\varepsilon_l/\mu_l}\cos\theta_l. \tag{47}$$

Четыре величины U_0, V_0, U и V, определяемые равенствами (46), связаны основным соотношением (32); следовательно,

$$A + R = (m'_{11} + m'_{12} p_l)\, T, \quad p_1(A - R) = (m'_{21} + m'_{22} p_l)\, T, \tag{48}$$

где m'_{ij} — элементы характеристической матрицы среды при $z = z_1$.

Из (48) мы найдем коэффициенты отражения и пропускания пленки

$$r = \frac{R}{A} = \frac{(m'_{11} + m'_{12} p_l)\, p_1 - (m'_{21} + m'_{22} p_l)}{(m'_{11} + m'_{12} p_l)\, p_1 + (m'_{21} + m'_{22} p_l)}, \tag{49}$$

$$t = \frac{T}{A} = \frac{2p_1}{(m'_{11} + m'_{12} p_l)\, p_1 + (m'_{21} + m'_{22} p_l)}. \tag{50}$$

*) Это подробно рассматривается в работе [25], стр. 611.

**) Мы рассматриваем амплитуды электрических векторов, когда изучаем волны TE-типа, а амплитуды магнитных векторов, когда изучаем волны TM-типа.

Отражательная и *пропускательная способности*, выраженные через r и t, имеют вид

$$\mathcal{R}=|r|^2, \qquad \mathcal{T}=\frac{p_l}{p_1}|t|^2. \tag{51}$$

Фазу δ_r величины r можно назвать *изменением фазы при отражении*, а фазу δ_t величины t — *изменением фазы при пропускании*. Изменение фазы δ_r отнесено к первой поверхности раздела, тогда как δ_t — к плоской границе между слоистой средой и последней полубесконечной средой.

Соответствующие формулы для волны TM-типа сразу же получаются из (49) — (51) путем замены величины p_1 и p_l на величины

$$q_1=\sqrt{\mu_1/\varepsilon_1}\cos\theta_1, \qquad q_l=\sqrt{\mu_l/\varepsilon_l}\cos\theta_l. \tag{52}$$

В этом случае r и t являются отношениями амплитуд магнитных, а не электрических векторов.

1.6.4. Однородная диэлектрическая пленка*). В оптике особый интерес представляют свойства однородной диэлектрической пленки, расположенной между двумя однородными средами. Поэтому этот случай мы рассмотрим более подробно. Все среды будем считать немагнитными ($\mu=1$).

Характеристическая матрица однородной диэлектрической пленки имеет вид (39). Обозначая через h толщину пленки и приписывая индексы 1, 2 и 3 величинам, которые относятся к трем средам (рис. 1.17), получим

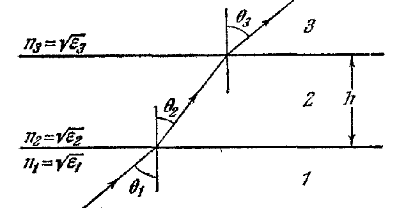


Рис. 1.17. Распространение электромагнитной волны через однородную пленку.

$$m'_{11}=m'_{22}=\cos\beta, \qquad m'_{12}=-\frac{i}{p_2}\sin\beta,$$
$$m'_{21}=-ip_2\sin\beta, \tag{53}$$

где

$$\beta=\frac{2\pi}{\lambda_0}n_2h\cos\theta_2$$

и

$$p_j=n_j\cos\theta_j, \qquad j=1,\ 2,\ 3. \tag{54}$$

Коэффициенты отражения и пропускания r и t можно получить, подставляя эти выражения в (49) и (50) и считая $l=3$. Окончательные формулы для них удобно записать, используя соответствующие коэффициенты r_{12}, t_{12} и r_{23}, t_{23}, связанные с отражением и пропусканием на первой и второй поверхностях раздела соответственно. Согласно формулам Френеля (1.5.20) и (1.5.21) получим для волны TE-типа

$$r_{12}=\frac{n_1\cos\theta_1-n_2\cos\theta_2}{n_1\cos\theta_1+n_2\cos\theta_2}=\frac{p_1-p_2}{p_1+p_2}, \tag{55}$$

$$t_{12}=\frac{2n_1\cos\theta_1}{n_1\cos\theta_1+n_2\cos\theta_2}=\frac{2p_1}{p_1+p_2}, \tag{56}$$

и аналогичные выражения для r_{23} и t_{23}. Если в формулы для r и t подставить эти величины, то они примут вид **)

$$r=\frac{r_{12}+r_{23}e^{2i\beta}}{1+r_{12}r_{23}e^{2i\beta}}, \tag{57}$$

$$t=\frac{t_{12}t_{23}e^{i\beta}}{1+r_{12}r_{23}e^{2i\beta}}. \tag{58}$$

---

*) Другой вывод основных формул, относящихся к свойствам одиночной диэлектрической пленки, встретится в п. 7.6.1. Эти формулы, разумеется, можно вывести непосредственно, использовав граничные условия, приведенные в п. 1.1.3, на каждой границе пленки (см. [32] или [26]).

**) Впервые эти формулы были получены другим способом в работе Эйри [33]; см. также [34].

Поэтому отражательная и пропускательная способности запишутся следующим образом:

$$\mathcal{R} = |r|^2 = \frac{r_{12}^2 + r_{23}^2 + 2r_{12}r_{23}\cos 2\beta}{1 + r_{12}^2 r_{23}^2 + 2r_{12}r_{23}\cos 2\beta} \quad (59)$$

и

$$\mathcal{T} = \frac{p_3}{p_1}|t|^2 = \frac{n_3\cos\theta_3}{n_1\cos\theta_1} \cdot \frac{t_{12}^2\, t_{23}^2}{1 + r_{12}^2 r_{23}^2 + 2r_{12}r_{23}\cos 2\beta}. \quad (60)$$

Прямой расчет, как и следовало ожидать, дает

$$\mathcal{R} + \mathcal{T} = 1.$$

Из (57) и (58) легко также рассчитать изменения фазы. Они определяются выражениями

$$\operatorname{tg}\delta_r = \operatorname{tg}(\arg r) = \frac{r_{23}(1 - r_{12}^2)\sin 2\beta}{r_{12}(1 + r_{23}^2) + r_{23}(1 + r_{12}^2)\cos 2\beta}, \quad (61)$$

$$\operatorname{tg}\delta_t = \operatorname{tg}(\arg t) = \frac{1 - r_{12}r_{23}}{1 + r_{12}r_{23}}\operatorname{tg}\beta. \quad (62)$$

Рассмотрим кратко некоторые выводы из этих формул. Во-первых, заметим, что выражения (59) и (60) остаются неизменными при замене β на $\beta + \pi$, т. е. при замене h на $h + \Delta h$, где

$$\Delta h = \frac{\lambda_0}{2n_2\cos\theta_2}. \quad (63)$$

Следовательно, *диэлектрические пленки, которые различаются по толщине на величину, кратную* $\lambda_0/2n_2\cos\theta_2$, *обладают одинаковыми отражательными и пропускательными способностями.*

Определим теперь оптическую толщину, для которой коэффициент отражения достигает максимума или минимума. Если положить

$$H = n_2 h, \quad (64)$$

то из (59) мы найдем

$$\frac{d\mathcal{R}}{dH} = 0, \quad \text{когда} \quad \sin 2\beta = 0,$$

т. е. когда

$$H = \frac{m\lambda_0}{4\cos\theta_2} \qquad (m = 0,\ 1,\ 2,\ \ldots).$$

Необходимо различать два случая:

1. *m нечетно*, т. е. H принимает значения

$$H = \frac{\lambda_0}{4\cos\theta_2},\quad \frac{3\lambda_0}{4\cos\theta_2},\quad \frac{5\lambda_0}{4\cos\theta_2},\ \ldots;$$

тогда $\cos 2\beta = -1$ и (59) переходит в

$$\mathcal{R} = \left(\frac{r_{12} - r_{23}}{1 - r_{12}r_{23}}\right)^2. \quad (65)$$

В частности, для *нормального падения* мы получим из (55)

$$r_{12} = \frac{n_1 - n_2}{n_1 + n_2}, \qquad r_{23} = \frac{n_2 - n_3}{n_2 + n_3}, \quad (66)$$

и (65) примет вид

$$\mathcal{R} = \left(\frac{n_1 n_3 - n_2^2}{n_1 n_3 + n_2^2}\right)^2. \quad (67)$$

2. *m четно*, т. е. оптическая толщина принимает значения

$$H = \frac{\lambda_0}{2\cos\theta_2},\quad \frac{2\lambda_0}{2\cos\theta_2},\quad \frac{3\lambda_0}{2\cos\theta_2},\ \ldots,$$

тогда $\cos 2\beta = 1$ и (59) переходит в

$$\mathcal{R} = \left(\frac{r_{12} + r_{23}}{1 + r_{12}r_{23}}\right)^2. \quad (68)$$

В частности, для *нормального падения* получим

$$\mathcal{R}=\left(\frac{n_1-n_3}{n_1+n_3}\right)^2, \tag{69}$$

и, как мы видим, $\mathcal{R}$ не зависит от n_2. Единственное отличие в случае наклонного падения состоит в том, что во всех формулах n_j заменяется на $n_j\cos\theta_j$ ($j=1$, 2, 3). Следовательно, *пластинка с оптической толщиной* $m\lambda_0/2\cos\theta_2$, *где* $m=1, 2, 3, \ldots$, *не влияет на интенсивность отраженного (или прошедшего) света.*

Нам нужно определить характер этих экстремальных значений. После простого расчета мы найдем, что при $H=m\lambda_0/4\cos\theta_2$ ($m=1, 2, \ldots$)

$$\left.\begin{aligned}&\frac{d^2\mathcal{R}}{dH^2}\gtrless 0\\ &\text{в соответствии с}\\ &(-1)^m r_{12}r_{23}(1+r_{12}^2r_{23}^2-r_{12}^2-r_{23}^2)\lessgtr 0.\end{aligned}\right\} \tag{70}$$

Верхний знак соответствует минимуму, нижний — максимуму. В частности, при *нормальном падении* r_{12} и r_{23} определяются выражениями (66), и мы имеем

$$\left.\begin{aligned}&\textit{максимум}, \text{ если } (-1)^m(n_1-n_2)(n_2-n_3)>0,\\ &\textit{минимум}, \text{ если } (-1)^m(n_1-n_2)(n_2-n_3)<0.\end{aligned}\right\} \tag{71}$$

Обычно первой средой является воздух ($n_1\sim 1$), и мы видим, что *для пленки с оптической толщиной* $\lambda_0/4$, $3\lambda_0/4$, $5\lambda_0/4$, ... *отражательная способность достигает максимума или минимума в зависимости от того, превышает ли показатель преломления пленки показатель преломления последней среды или он меньше его. Для пленки, с оптической толщиной* $\lambda_0/2$, $2\lambda_0/2$, $3\lambda_0/2$, ... *справедливо обратное.*

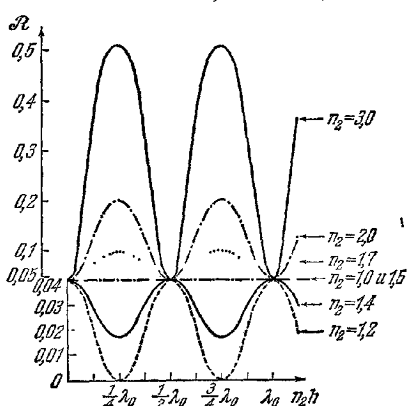


Рис. 1.18. Отражательная способность диэлектрической пленки с показателем преломления n_2 как функция ее оптической толщины [36]. $\theta_1=0$, $n_1=1$, $n_3=1,5$.

Эти результаты, которые иллюстрируются рис. 1.18, оказались в хорошем согласии с экспериментом (см., например, [35]).

Из предыдущего анализа ясно, что пластинку с достаточно низким показателем преломления, оптическая толщина которой составляет четверть длины волны, можно использовать в качестве *просветляющей пленки*, т. е. пленки, уменьшающей отражательную способность поверхности. Чаще всего для этой цели применяются два вещества: криолит ($n\sim 1,35$) и фтористый магний (MgF_2, $n\sim 1,38$).

Согласно (67) отражательная способность при нормальном падении была бы точно равна нулю, если бы

$$n_2=\sqrt{n_1n_3}. \tag{72}$$

При $n_1=1$, $n_3=1,5$ это требует, чтобы n_2 примерно равнялось 1,22, условие, которому практически нельзя удовлетворить. Однако более полный анализ выражения (59) показывает, что при наклонном падении можно получить равную нулю отражательную способность для волны *TM*-типа (электрический вектор параллелен плоскости падения), но не для волны *TE*-типа (электрический вектор перпендикулярен плоскости падения), т. е. при благоприятных условиях

можно получить одновременно $\mathcal{R}_{\parallel}=0$ и $\mathcal{R}_{\perp}\neq 0$. Следовательно, тонкая пленка из подходящего диэлектрика способна служить *поляризатором*, работающим на отражение. Такой поляризатор можно считать более общим вариантом простого устройства, обсуждавшегося ранее в связи с углом Брюстера. Чтобы получить большое значение для $\mathcal{R}_{\perp}$ (при $\mathcal{R}_{\parallel}=0$) показатель преломления пленки n_2 должен быть по возможности большим [37]. Например, при $n_1=1$, $n_2=2{,}5$, $n_3=1{,}53$ мы получим $\mathcal{R}_{\parallel}=0$, $\mathcal{R}_{\perp}=0{,}79$, когда $\theta_1=$ $=74°30'$.

Если стеклянную поверхность покрыть материалом с достаточно высоким показателем преломления, то отражательная способность поверхности, согласно предыдущему анализу, сильно увеличится (см. рис. 1.18 и 1.19). Такая поверхность будет действовать как хороший делитель пучка. Для этой цели очень подходят покрытия из двуокиси титана (TiO_2, $n\sim 2{,}45$) или сернистого цинка (ZnS, $n\sim 2{,}3$), которые дают максимальную отражательную способность, примерно равную 0,3. Существуют и другие вещества, обладающие высокими показателями преломления, но они поглощают часть падающего света. Например, пользуясь покрытием из стибнита (Sb_2S_3, $n\sim 2{,}8$), можно достичь значений $\mathcal{R}=\mathcal{T}=0{,}46$, но при этом 8% падающего света поглощается пленкой.

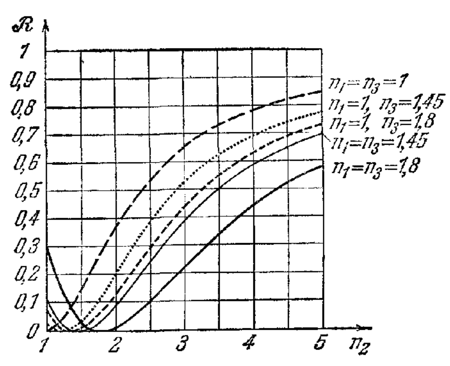


Рис. 1.19. Отражательная способность четвертьволновой пленки ($n_2h=\lambda_0/4$) при нормальном падении как функция показателя преломления пленки n_2 [35].

Интересно также исследовать случай полного внутреннего отражения на первой поверхности раздела. При этом получим

$$n_1\sin\theta_1>n_2,\qquad n_1\sin\theta_1<n_3$$

и (см. уравнение (1.5.54))

$$n_2\cos\theta_2=-i\sqrt{n_1^2\sin^2\theta_1-n_2^2}. \tag{73}$$

Коэффициенты отражения для обеих поверхностей раздела теперь будут равны (см. (1.5.21))

$$\left.\begin{aligned} r_{12}&=\frac{n_1\cos\theta_1-i\sqrt{n_1^2\sin^2\theta_1-n_2^2}}{n_1\cos\theta_1+i\sqrt{n_1^2\sin^2\theta_1-n_2^2}},\\ r_{23}&=\frac{-i\sqrt{n_1^2\sin^2\theta_1-n_2^2}-n_3\cos\theta_3}{-i\sqrt{n_1^2\sin^2\theta_1-n_2^2}+n_3\cos\theta_3}.\end{aligned}\right\} \tag{74}$$

Если положить

$$k_0n_2h\cos\theta_2=ib, \tag{75}$$

где, согласно (73),

$$b=-\frac{2\pi}{\lambda_0}h\sqrt{n_1^2\sin^2\theta_1-n_2^2}, \tag{76}$$

то для коэффициента отражения мы получим вместо (57) выражение

$$r=\frac{r_{12}+r_{23}e^{-2b}}{1+r_{12}r_{23}e^{-2b}}. \tag{77}$$

Так как $|r_{12}|=|r_{23}|=1$, то r_{12} и r_{23} имеют вид

$$r_{12}=\exp(i\varphi_{12}),\qquad r_{23}=\exp(i\varphi_{23}), \tag{78}$$

где величины φ вещественны. Следовательно, отражательная способность равна

$$\mathcal{R}=|r|^2=\frac{e^{2b}+e^{-2b}+2\cos(\varphi_{12}-\varphi_{23})}{e^{2b}+e^{-2b}+2\cos(\varphi_{12}+\varphi_{23})}. \tag{79}$$

В отличие от предыдущего случая, $\mathcal{R}$ уже не является периодической функцией толщины пленки. Выражение (76) показывает, что если пренебречь зависимостью показателя преломления от длины волны, то величина b обратно пропорциональна длине волны. Так как для достаточно больших значений b величина $\mathcal{R}$ практически равна единице, то более короткие длины волн не будут пропускаться. В этом случае пленка действует как фильтр нижних частот, т. е. как фильтр, пропускающий лишь длинные волны.

Мы видели, что с помощью диэлектрических пленок из подходящего материала можно получить много полезных эффектов. Из дальнейшего станет ясно, что, применяя систему из нескольких таких пленок, можно еще больше усилить нужные свойства. Характеристическую матрицу подобной *многослойной системы* можно получить с помощью теоремы, выраженной в виде (41) *). Мы рассмотрим подробно лишь случай периодической многослойной системы.

1.6.5. Периодические слоистые среды. Слоистая периодическая среда с периодом h характеризуется диэлектрической проницаемостью ε и магнитной проницаемостью μ, зависящими только от z; они имеют вид

$$\varepsilon(z+jh)=\varepsilon(z), \qquad \mu(z+jh)=\mu(z),$$

где j — любое целое число из некоторого фиксированного интервала $1\leqslant j\leqslant N$.

Пусть $M(h)$ — характеристическая матрица, соответствующая одному периоду. Запишем **)

$$M(h)=\begin{bmatrix} m_{11} & m_{12}\\ m_{21} & m_{22}\end{bmatrix}. \tag{80}$$

В этом случае, в соответствии с (41) и с учетом периодичности среды, получим

$$M(Nh)=\underbrace{M(h)\,M(h)\,\ldots\,M(h)}_{N\text{ раз}}=[M(h)]^N. \tag{81}$$

Для определения элементов матрицы $M(Nh)$ воспользуемся результатом, полученным в теории матриц, согласно которому N-я степень унимодулярной матрицы $M(h)$ равна ***)

$$[M(h)]^N=\begin{bmatrix} m_{11}\mathcal{U}_{N-1}(a)-\mathcal{U}_{N-2}(a) & m_{12}\mathcal{U}_{N-1}(a)\\ m_{21}\mathcal{U}_{N-1}(a) & m_{22}\mathcal{U}_{N-1}(a)-\mathcal{U}_{N-2}(a)\end{bmatrix}, \tag{82}$$

где

$$a=\frac{1}{2}(m_{11}+m_{22}), \tag{83}$$

а $\mathcal{U}_N$ — *полиномы Чебышева* второго рода:

$$\mathcal{U}_N(x)=\frac{\sin[(N+1)\arccos x]}{\sqrt{1-x^2}}. \tag{84}$$

---

*) Формулы, относящиеся к многослойным системам, были получены многими авторами, см., например, [38, 39].

**) Здесь мы опускаем штрих у элементов матрицы.

***) В правильности этого результата можно убедиться методом индукции, используя рекуррентное соотношение

$$\mathcal{U}_j(x)=2x\mathcal{U}_{j-1}(x)-\mathcal{U}_{j-2}(x),$$

которое следует как тождество из определения полиномов Чебышева. Прямое доказательство, основанное на теории матриц, приведено в работе [25].

Эти полиномы *) удовлетворяют следующим условиям ортогональности и нормировки:

$$\int_{-1}^{+1} \mathcal{U}_m(x)\,\mathcal{U}_n(x)\sqrt{1-x^2}\,dx = \begin{cases} 0 & \text{при} \quad n \neq m, \\ \dfrac{\pi}{2} & \text{при} \quad n = m. \end{cases}$$

Для удобства выпишем в явном виде первые шесть полиномов:

$$\mathcal{U}_0(x) = 1, \qquad \mathcal{U}_3(x) = 8x^3 - 4x,$$
$$\mathcal{U}_1(x) = 2x, \qquad \mathcal{U}_4(x) = 16x^4 - 12x^2 + 1,$$
$$\mathcal{U}_2(x) = 4x^2 - 1, \quad \mathcal{U}_5(x) = 32x^5 - 32x^3 + 6x.$$

Многослойная система обычно состоит из чередующихся однородных слоев с низким и высоким показателями преломления n_2 и n_3 и толщинами

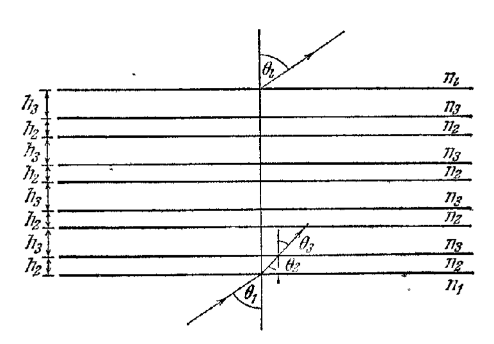


Рис. 1.20. Периодическая многослойная система.

h_2 и h_3, расположенных между двумя однородными средами с показателями преломления n_1 и n_l (рис. 1.20). Мы снова считаем среду немагнитной ($\mu = 1$) и полагаем

$$\left.\begin{aligned} \beta_2 &= \frac{2\pi}{\lambda_0} n_2 h_2 \cos\theta_2, & \beta_3 &= \frac{2\pi}{\lambda_0} n_3 h_3 \cos\theta_3, \\ p_2 &= n_2 \cos\theta_2, & p_3 &= n_3 \cos\theta_3, \\ & & h = h_2 + h_3. & \end{aligned}\right\} \tag{85}$$

Согласно (39) и (41) в этом случае характеристическая матрица $M_2(h)$ одного периода имеет вид

$$M_2(h) = \begin{bmatrix} \cos\beta_2 & -\dfrac{i}{p_2}\sin\beta_2 \\ -ip_2\sin\beta_2 & \cos\beta_2 \end{bmatrix} \begin{bmatrix} \cos\beta_3 & -\dfrac{i}{p_3}\sin\beta_3 \\ -ip_3\sin\beta_3 & \cos\beta_3 \end{bmatrix} =$$
$$= \begin{bmatrix} \cos\beta_2\cos\beta_3 - \dfrac{p_3}{p_2}\sin\beta_2\sin\beta_3 & -\dfrac{i}{p_3}\cos\beta_2\sin\beta_3 - \dfrac{i}{p_2}\sin\beta_2\cos\beta_3 \\ -ip_2\sin\beta_2\cos\beta_3 - ip_3\cos\beta_2\sin\beta_3 & \cos\beta_2\cos\beta_3 - \dfrac{p_2}{p_3}\sin\beta_2\sin\beta_3 \end{bmatrix}. \tag{86}$$

Следовательно, в соответствии с (81) характеристическая матрица $M_{2N}(Nh)$ многослойной системы (с полным числом пленок $2N$) определяется следующими формулами, полученными Абеле:

$$M_{2N}(Nh) = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{bmatrix}, \tag{87}$$

*) Таблицы полиномов Чебышева были опубликованы Национальным Бюро Стандартов [40]; там же приведена сводка их наиболее важных свойств (см. также [41]).

где

$$\left.\begin{aligned}\mathcal{M}_{11}&=\left(\cos\beta_2\cos\beta_3-\frac{p_3}{p_2}\sin\beta_2\sin\beta_3\right)\mathcal{U}_{N-1}(a)-\mathcal{U}_{N-2}(a),\\ \mathcal{M}_{12}&=-i\left(\frac{1}{p_3}\cos\beta_2\sin\beta_3+\frac{1}{p_2}\sin\beta_2\cos\beta_3\right)\mathcal{U}_{N-1}(a),\\ \mathcal{M}_{21}&=-i\,(p_2\sin\beta_2\cos\beta_3+p_3\cos\beta_2\sin\beta_3)\,\mathcal{U}_{N-1}(a),\\ \mathcal{M}_{22}&=\left(\cos\beta_2\cos\beta_3-\frac{p_2}{p_3}\sin\beta_2\sin\beta_3\right)\mathcal{U}_{N-1}(a)-\mathcal{U}_{N-2}(a)\end{aligned}\right\}\tag{88}$$

и

$$a=\cos\beta_2\cos\beta_3-\frac{1}{2}\left(\frac{p_2}{p_3}+\frac{p_3}{p_2}\right)\sin\beta_2\sin\beta_3.\tag{89}$$

Коэффициенты отражения и пропускания многослойной системы сразу же получаются при подстановке этих выражений в (49) и (50).

Особый интерес представляет случай, когда два основных слоя имеют одну и ту же оптическую толщину (обычно $\lambda_0/4$), т. е. когда

$$n_2h_2=n_3h_3,\tag{90}$$

и падение нормально ($\theta_1=0$). Тогда

$$\beta_2=\beta_3=\frac{2\pi}{\lambda_0}n_2h_2=\frac{2\pi}{\lambda_0}n_3h_3;\tag{91}$$

если эту общую для обоих слоев величину обозначить через β, то аргумент полиномов Чебышева примет вид

$$a=\cos^2\beta-\frac{1}{2}\left(\frac{n_2}{n_3}+\frac{n_3}{n_2}\right)\sin^2\beta.\tag{92}$$

Видно, что величина a не может превышать единицу, но для некоторых значений β она может стать меньше -1. В этом случае $\arccos a$ — мнимая величина, а так как для любого χ

$$\sin i\chi=i\,\mathrm{sh}\,\chi=i\,\frac{e^{\chi}-e^{-\chi}}{2},$$

то, следовательно, $\mathcal{U}_N$ будет вести себя как экспонента. Отсюда следует, что отражательная способность такой многослойной системы быстро увеличивается с ростом числа периодов.

Для *четвертьволновых пленок* ($n_2h_2=n_3h_3=\lambda_0/4$) при нормальном падении (вновь считая среды немагнитными) имеем

$$\beta=\pi/2,\qquad p_2=n_2,\qquad p_3=n_3,\tag{93}$$

и (86) переходит в

$$\mathsf{M}_2(h)=\begin{bmatrix}-\dfrac{n_3}{n_2} & 0\\ 0 & -\dfrac{n_2}{n_3}\end{bmatrix}.\tag{94}$$

Характеристическая матрица (87) многослойной системы, основным периодом которой служит такой двойной слой, имеет вид

$$\mathsf{M}_{2N}(Nh)=\begin{bmatrix}\left(-\dfrac{n_3}{n_2}\right)^N & 0\\ 0 & \left(-\dfrac{n_2}{n_3}\right)^N\end{bmatrix},\tag{95}$$

что можно подтвердить, умножая матрицу (94) саму на себя N раз. Согласно (49) и (50) отражательная способность рассматриваемой системы равна

$$\mathcal{R}_{2N}=\left(\frac{1-\dfrac{n_l}{n_1}\left(\dfrac{n_2}{n_3}\right)^{2N}}{1+\dfrac{n_l}{n_1}\left(\dfrac{n_2}{n_3}\right)^{2N}}\right)^2.\tag{96}$$

Отсюда следует, что для фиксированного числа двойных слоев *) N отражательная способность $\mathcal{R}_{2N}$ увеличивается при увеличении n_2/n_3, а для фиксированного значения этого отношения $\mathcal{R}_{2N}$ увеличивается с ростом N.

Иногда, например, при покрытии пластин интерферометра Фабри — Перо (см. § 7.6) слои располагают так, что их показатели преломления образуют следующую последовательность: $n_2, n_3, n_2, n_3, \ldots, n_2, n_3, n_2$. Характеристическая матрица такой многослойной системы равна

$$\mathsf{M}_{2N+1} = \mathsf{M}_{2N} \cdot \mathsf{M}, \qquad (97)$$

где M_{2N} определяется выражением (87), а M — характеристическая матрица последней пленки этого ряда. Для четвертьволновых пленок при нормальном падении M_{2N} переходит в (95), $\beta_2 = \pi/2$, и (97) примет вид

$$\mathsf{M}_{2N+1} = \begin{bmatrix} 0 & -\frac{i}{n_2}\left(-\frac{n_3}{n_2}\right)^N \\ -in_2\left(-\frac{n_2}{n_3}\right)^N & 0 \end{bmatrix}. \qquad (98)$$

Подстановка в (49) и (50) даст отражательную способность

$$\mathcal{R}_{2N+1} = \left\{\frac{1 - \frac{n_2}{n_1}\frac{n_2}{n_l}\left(\frac{n_2}{n_3}\right)^{2N}}{1 + \frac{n_2}{n_1}\frac{n_2}{n_l}\left(\frac{n_2}{n_3}\right)^{2N}}\right\}^2. \qquad (99)$$

Таблица 1.3

Отражательная способность $\mathcal{R}_{2N+1}$ многослойных систем, состоящих из периодической совокупности четвертьволновых пленок сернистого цинка и криолита, при нормальном падении ($n_1 = 1$, $n_2 = 2{,}3$, $n_3 = 1{,}35$, $n_l = 1{,}52$, $n_2h_2 = n_3h_3 = \lambda_0/4$, $\lambda_0 = 5460$ Å, $\theta_1 = 0$).

N	$\mathcal{R}_{2N+1}$
0	0,306
1	0,676
2	0,872 (0,865)
3	0,954 (0,945)
4	0,984 (0,97)

В скобках приведены экспериментальные результаты, полученные в работе [43].

Как мы видим, отражательная способность быстро увеличивается с увеличением отношения n_2/n_3 и N (см. табл. 1.3).

ЛИТЕРАТУРА

1. P. A. Franken, A. E. Hill, C. W. Peters, G. Weinrich, Phys. Rev. Lett. 7, 118 (1961).
2. J. A. Armstrong, N. Bloembergen, J. Ducuing, P. S. Pershan, Phys. Rev. 127, 1918 (1962).
3. N. Bloembergen, P. S. Pershan, Phys. Rev. 128, 606 (1962).
4. N. Bloembergen, Nonlinear Optics, W. A. Benjamin, Inc., New York, 1965; (Н. Бломберген, Нелинейная оптика, «Мир», 1966.) P. S. Pershan, в сб.: «Progr. in Optics», ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1965, p. 83; P. A. Eranken, J. E. Ward, Rev. Mod. Phys. 35, 23 (1963).
5. A. Sommerfeld, Electrodynamics, Acad. Press. New York, 1952.
6. J. A. Stratton, Electromagnetic Theory, McGraw-Hill, New York, 1941. (Дж. А. Стрэттон, Теория электромагнетизма, ИЛ, 1948.)
7. W. C. Röntgen, Ann. d. Physik 35, 264 (1888); 40, 93 (1890).
8. E. Bergstrand, в кн. «Encyclopedia of Physics», ed. S. Flügge, Springer, Berlin, 1956, Vol. 24, p. 1.
9. R. T. Birge, в сб. «Rep. Progr. Phys.», The Phys. Soc. London, 1941, Vol. 8, p. 90.
10. L. Boltzmann, Wien. Ber. 69, 795 (1874); Pogg. Ann. 155, 403 (1875); Wiss. Abh. Physik-techn. Reichsanst. 1, № 26, 537.
11. A. Sommerfeld, Ann. d. Physik 44, 177 (1914).
12. L. Brillouin, Ann. d. Physik 44, 203 (1914).
13. L. Brillouin, Wave Propagation and Group Velocity, Acad. Press, New York, 1960, pp. 17, 43.

*) Свойства системы двойных слоев подробно рассматриваются в статье [42].

14. H. G. Baerwald, Ann. d. Physik 6, 295 (1930); 7, 731 (1930); 8, 565 (1931).
15. F. Borgnis, Z. Phys. 117, 642 (1941).
16. L. J. F. Broer, Appl. Sci. Res A2, 329 (1951).
17. H. Poincaré, Théorie Mathématique de la Lumière, Géorges Carré, Paris, 1892, Vol. 2, Ch. 12.
18. H. G. Jerrard, J. Opt. Soc. Amer. 44, 634 (1954).
19. M. J. Walker, Amer. J. Phys. 22, 170 (1954).
20. S. Pancharatnam, Proc. Indian Acad. Sci. A44, 247 (1956).
21. A. Nisbet, E. Wolf, Proc. Cambr. Phil. Soc. 50, 614 (1954).
22. О. Д. Хвольсон, Курс физики, РСФСР, ГИЗ, Берлин, 1923, т. 2, стр. 623.
23. W. Culshaw, D. S. Jones, Proc. Phys. Soc. B66, 859 (1954).
24. J. Picht, Ann. d. Physik, 3, 433 (1929).
25. F. Abelés, Ann. de Physique 5, 596—640, 706—782 (1950).
26. H. Mayer, Physik dunner Schichten, Wissenschaftliche Verlagsgesellschaft, Stuttgart, 1950.
27. S. Methfessel, Dünne Schichten, VEB Wilhelm Knapp Verlag, Halle (Saale), 1953. (С. Метфессель, Тонкие пленки, их изготовление и измерение, Госэнергоиздат, 1963.)
28. O. S. Heavens, Optical Properties of Thin Solid Films, Butterworths Sci. Publ., London, 1955.
29. R. B. Muchmore, J. Opt. Soc. Amer. 38, 20 (1948).
30. K. Schuster, Ann. d. Physik 4, 352 (1949); R. Kronig, R. S. Blaisse, J. J. v. v. Sande, J. Appl. Sci. Res. B1, 63 (1947).
31. R. Jacobsson, в сб.: «Progr. in Optics», ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1965, p. 247.
32. M. Born, Optik, Springer, Berlin, 1933. (М. Борн, Оптика, ГНТИУ, 1937.)
33. G. B. Airy, Phil. Mag. 2, 20 (1833).
34. G. B. Airy, Ann. Phys. a. Chem. 41, 512 (1837).
35. K. Hammer, Z. techn. Phys. 24, 169 (1943).
36. R. Messner, Zeiss Nachr. 4, (H9), 253 (1943).
37. H. Schröder, Optik 3, 499 (1948).
38. R. L. Mooney, J. Opt. Soc. Amer. 36, 256 (1946).
39. W. Weinstein, J. Opt. Soc. Amer. 37, 576 (1947).
40. Applied Mathematics Series 9, Nat. Bur. Stand. Washington, 1952.
41. Higher Transcendental Functions, Bateman Manuscript Project, McGraw-Hill, New York, 1953, Vol. 2, p. 183.
42. C. Dufour, A. Herpin, Rev. d'Opt. 32, 321 (1953).
43. P. Giacomo, C. R. Acad. Sci. 235, 1627 (1952).
44*. Г. Д. Бурдун, Единицы физических величин, Стандартгиз, 1962, стр. 29.
45*. С. А. Ахманов, Р. В. Хохлов, Проблемы нелинейной оптики (Электромагнитные волны в нелинейных диспергирующих средах), 1962—1963. Ин-т научной информации АН СССР, 1964.
46*. В. М. Файн, Я. И. Ханин, Квантовая радиофизика, «Сов. радио», 1965.
47*. Л. И. Мандельштам, Phys. Z. 15, 220 (1914); Полное собрание трудов, изд. АН СССР, 1948, т. 1, стр. 261.

ГЛАВА 2

ЭЛЕКТРОМАГНИТНЫЕ ПОТЕНЦИАЛЫ И ПОЛЯРИЗАЦИЯ

В предыдущей главе влияние вещества на электромагнитное поле выражалось через ряд макроскопических постоянных. Однако область их применения ограничена и фактически, пользуясь ими, нельзя адекватно описать такие процессы, как излучение, поглощение и дисперсия света. Полное описание этих явлений потребовало бы подробного изучения атомистики и поэтому выходит за рамки настоящей книги.

Можно, однако, адекватно описать взаимодействие поля и вещества при помощи простой модели, вполне применимой для большинства разделов оптики. Для этой цели каждый из векторов **D** и **B** выражают в виде суммы двух членов *). Один из них считают равным соответствующему вектору поля в вакууме и полагают, что другой описывает влияние среды. Таким образом, мы приходим к необходимости ввести два новых вектора для описания влияния вещества: *электрическую поляризацию* **P** и *магнитную поляризацию*, или *намагничение* **M**. Вместо материальных уравнений (1.1.10) и (1.1.11), связывающих **D** и **B** с **E** и **H**, теперь появятся уравнения, связывающие **P** и **M** с **E** и **H**. Эти новые уравнения имеют более прямой физический смысл и приводят к следующей концепции распространения электромагнитного поля в среде.

Электромагнитное поле создает в данном элементе объема известные степени поляризации **P** и **M**, которые в первом приближении пропорциональны полю, причем коэффициент пропорциональности служит мерой реакции поля. Тогда каждый элемент объема становится источником новой вторичной, или рассеянной, волны, амплитуда которой простым образом связана с **P** и **M**. Все вторичные волны комбинируются друг с другом и с падающим полем и образуют полное поле, причем именно оно и считается основным. Формализуя сказанное выше, мы получим два интегральных уравнения **), которые, как легко показать, эквивалентны дифференциальным уравнениям Максвелла, но описывают распространение электромагнитного поля способом, более ясно связанным с атомным строением вещества.

Из теории будут получены следующие два основных результата: 1) формула Лоренц — Лоренца ***), которая связывает макроскопические оптические свойства среды с числом и свойствами рассеивающих частиц, и 2) так называемая теорема погашения Эвальда и Озеена, которая показывает, каким образом внешнее электромагнитное возмущение, распространяющееся со скоростью света в вакууме, точно компенсируется и заменяется в веществе вторичным возмущением, распространяющимся с соответственно меньшей скоростью.

---

*) Согласно замечанию на стр. 24 правильнее было бы писать **H** вместо **B**. Отступление от общей теории делается здесь просто ради удобства иметь положительные знаки в правой части уравнения (2.2.2), как это обычно принято.

**) В случае немагнитных веществ остается лишь одно интегральное уравнение. Второе уравнение переходит в относительно простое выражение для магнитного поля, и его можно найти, когда получено решение первого уравнения. Рассматриваемые здесь «материальные» интегральные уравнения необходимо отличать от «геометрических» интегральных уравнений, которые применяются при изучении некоторых дифракционных проблем (см. гл. 11).

***) В русской литературе фамилии обоих ученых (Lorentz и Lorenz) обычно транскрибируются одинаковым образом — Лоренц. Однако поскольку в тексте встречаются ссылки на каждого из авторов в отдельности, мы сочли целесообразным принять тоже встречающуюся в ряде случаев разную транскрипцию. (*Прим. ред.*)

6*

Теория дает и другой математический подход к рассмотрению некоторых проблем электромагнитной теории. Мы проиллюстрируем его выводом законов преломления и отражения и формул Френеля, а затем решением более сложной проблемы в гл. 12.

Вывод этих результатов потребует несколько другого математического аппарата. Поэтому вначале мы рассмотрим представление электромагнитного поля через так называемые запаздывающие потенциалы, которые являются обобщениями хорошо известных статических потенциалов. Выражение потенциалов через векторы поляризации приводит к необходимости ввести другие вспомогательные величины, известные как векторы Герца. В §§ 2.1 и 2.2 мы рассмотрим математические предпосылки, а в § 2.3 кратко изложим исходные физические понятия. Интегральные уравнения и упоминавшиеся выше две основные теоремы будут выведены в § 2.4.

§ 2.1. Электродинамические потенциалы в вакууме

2.1.1. Векторные и скалярные потенциалы. Рассмотрим электромагнитное поле в вакууме, обусловленное заданным распределением зарядов ρ ($\mathbf{r}$, t) и токов $\mathbf{j}$ ($\mathbf{r}$, t). Оно удовлетворяет уравнениям Максвелла (1.1.1)—(1.1.4). В вакууме $\mathbf{D}=\mathbf{E}$ и $\mathbf{B}=\mathbf{H}$, и эти уравнения можно переписать в виде

$$\operatorname{rot}\mathbf{B}-\frac{1}{c}\dot{\mathbf{E}}=\frac{4\pi}{c}\mathbf{j}, \tag{1}$$

$$\operatorname{rot}\mathbf{E}+\frac{1}{c}\dot{\mathbf{B}}=0, \tag{2}$$

$$\operatorname{div}\mathbf{E}=4\pi\rho, \tag{3}$$

$$\operatorname{div}\mathbf{B}=0. \tag{4}$$

Так как дивергенция ротора любого вектора равна нулю, то уравнение (4) будет удовлетворяться, если положить

$$\mathbf{B}=\operatorname{rot}\mathbf{A}, \tag{5}$$

где $\mathbf{A}$ — произвольная векторная функция координат и времени. Если соотношение (5) подставить во второе уравнение Максвелла, мы получим

$$\operatorname{rot}\left(\mathbf{E}+\frac{1}{c}\dot{\mathbf{A}}\right)=0. \tag{6}$$

Уравнение (6) будет удовлетворяться, если

$$\mathbf{E}=-\frac{1}{c}\dot{\mathbf{A}}-\operatorname{grad}\varphi; \tag{7}$$

здесь φ — произвольная скалярная функция. Величины $\mathbf{A}$ и φ необходимо определить таким образом, чтобы удовлетворить оставшимся уравнениям Максвелла.

Подставляя (5) и (7) в (1) и (3) и используя тождества

$$\operatorname{rot}\operatorname{rot}\equiv\operatorname{grad}\operatorname{div}-\nabla^2 \quad \text{и} \quad \operatorname{div}\operatorname{grad}\equiv\nabla^2,$$

получим

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\ddot{\mathbf{A}}-\operatorname{grad}\left(\operatorname{div}\mathbf{A}+\frac{1}{c}\dot{\varphi}\right)=-\frac{4\pi}{c}\mathbf{j} \tag{8}$$

и

$$\nabla^2\varphi-\frac{1}{c^2}\ddot{\varphi}+\frac{1}{c}\frac{\partial}{\partial t}\left(\operatorname{div}\mathbf{A}+\frac{1}{c}\dot{\varphi}\right)=-4\pi\rho. \tag{9}$$

Если связать $\mathbf{A}$ и φ соотношением

$$\operatorname{div}\mathbf{A}+\frac{1}{c}\dot{\varphi}=0, \tag{10}$$

то (8) и (9) перейдут в неоднородные волновые уравнения

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\ddot{\mathbf{A}}=-\frac{4\pi}{c}\mathbf{j} \qquad (11)$$

и

$$\nabla^2\varphi-\frac{1}{c^2}\ddot{\varphi}=-4\pi\rho. \qquad (12)$$

Функции $\mathbf{A}$ и φ, из которых посредством соотношений (5) и (7) можно определить $\mathbf{B}$ и $\mathbf{E}$, известны как *магнитный векторный потенциал* и *электрический скалярный потенциал* соответственно. Соотношение (10), связывающее оба потенциала, называется *условием Лоренца.* Заметим, что оно согласуется с уравнением непрерывности (1.1.5)

$$\dot{\rho}+\operatorname{div}\mathbf{j}=0. \qquad (13)$$

Легко показать, что выражения (11), (12) и (10) не определяют потенциалы однозначно. Действительно, если мы добавим к $\mathbf{A}$ вектор $\operatorname{grad}\chi$, где χ произвольно, то $\mathbf{B}$ не изменится, а если, кроме того, φ заменить на $\varphi-\dot{\chi}/c$, то $\mathbf{E}$ также не изменится. Другими словами, $\mathbf{B}$ и $\mathbf{E}$ инвариантны при преобразовании

$$\mathbf{A}'=\mathbf{A}+\operatorname{grad}\chi, \qquad (14а)$$

$$\varphi'=\varphi-\frac{1}{c}\dot{\chi}. \qquad (14б)$$

Из (10) и (14) получим

$$\operatorname{div}\mathbf{A}'+\frac{1}{c}\dot{\varphi}'+\left(\nabla^2\chi-\frac{1}{c^2}\ddot{\chi}\right)=0. \qquad (15)$$

Следовательно, $\mathbf{A}'$ и φ' будут удовлетворять соотношению Лоренца, если на χ наложить условие

$$\nabla^2\chi-\frac{1}{c^2}\ddot{\chi}=0. \qquad (16)$$

Уравнения (14), подчиняющиеся условию (16), выражают так называемое *калибровочное преобразование.* Калибровочное преобразование можно использовать для упрощения записи векторов поля. Например, в области с нулевой плотностью заряда ρ величина φ удовлетворяет однородному волновому уравнению

$$\nabla^2\varphi-\frac{1}{c^2}\ddot{\varphi}=0; \qquad (17)$$

тогда величину χ можно выбрать так, чтобы скалярный потенциал оказался равным нулю. Согласно (14б) и (16) для этого необходимо положить

$$\chi=c\int\varphi\,dt. \qquad (18)$$

Поле можно найти с помощью одного лишь векторного потенциала посредством соотношений *) (штрих над $\mathbf{A}$ опущен)

$$\mathbf{B}=\operatorname{rot}\mathbf{A}, \qquad \mathbf{E}=-\frac{1}{c}\dot{\mathbf{A}}, \qquad (19)$$

тогда как условие Лоренца переходит в

$$\operatorname{div}\mathbf{A}=0. \qquad (20)$$

2.1.2. Запаздывающие потенциалы. Рассмотрим решения неоднородных волновых уравнений (11) и (12) для векторного и скалярного потенциалов, подчиняющиеся соотношению (10), и покажем вначале, что этим уравнениям

*) Уравнение (19) описывает поле в свободной от зарядов области в вакууме тремя скалярными волновыми функциями (декартовы компоненты вектора $\mathbf{A}$). Из-за соотношения (20), их, однако, нельзя считать независимыми. Нетрудно показать, что фактически в такой области поле определяется двумя вещественными скалярными волновыми функциями (см., например, [1—3]).

удовлетворяют следующие функции:

$$\mathbf{A}(\mathbf{r},\ t)=\frac{1}{c}\int\frac{\mathbf{j}(\mathbf{r}',\ t-R/c)}{R}dV',\tag{21}$$

$$\varphi(\mathbf{r},\ t)=\int\frac{\rho(\mathbf{r}',\ t-R/c)}{R}dV'.\tag{22}$$

Здесь

$$R=|\mathbf{r}-\mathbf{r}'|=\sqrt{(x-x')^2+(y-y')^2+(z-z')^2}\tag{23}$$

— расстояние между точкой $\mathbf{r}$ $(x,\ y,\ z)$ и точкой $\mathbf{r}'$ $(x',\ y',\ z')$ элемента объема dV'. Интегрирование проводится по всему пространству.

Чтобы убедиться в том, что (22) удовлетворяет неоднородному волновому уравнению для скалярного потенциала, мы вообразим, что точка $\mathbf{r}$ окружена сферой радиуса a с центром в этой точке, и разделим (22) на две части

$$\varphi=\varphi_1+\varphi_2,\tag{24}$$

где φ_1— вклад в интеграл от внутренней части сферы, а φ_2— вклад от остальной части пространства. Так как $R\neq 0$ для каждой точки $\mathbf{r}'$ $(x',\ y',\ z')$ вне сферы, то φ_2 можно продифференцировать под знаком интеграла. Тогда прямым расчетом легко убедиться, что φ_2 удовлетворяет однородному волновому уравнению

$$\nabla^2\varphi_2-\frac{1}{c^2}\ddot{\varphi}_2=0.\tag{25}$$

При этом ясно, что φ_2 представляет суперпозицию ряда сферических волн (см. уравнение (1.3.12)). В случае φ_1 мы должны, однако, поступить иначе, так как подынтегральное выражение имеет особенность в центре сферы $R=0$.

Заметим, что, выбрав радиус сферы достаточно малым, можно добиться того (при условии, что ρ — непрерывная функция $\mathbf{r}$ и t), чтобы для всех точек $\mathbf{r}'$ внутри сферы $\rho(\mathbf{r}',\ t-R/c)$ отличалось от $\rho(\mathbf{r},\ t)$ на величину, меньшую любого наперед заданного значения. Следовательно, когда радиус a стремится к нулю, $\nabla^2\varphi_1$ будет все более и более приближаться к значению, соответствующему электростатическому потенциалу равномерно заряженной сферы с плотностью заряда ρ; т. е. для достаточно малых a имеем

$$\nabla^2\varphi_1=-4\pi\rho(\mathbf{r},\ t).\tag{26}$$

Аналогично $\ddot{\varphi}_1\to 0$, когда $a\to 0$. В самом деле, если a достаточно мало, мы можем написать

$$\ddot{\varphi}_1=\ddot{\rho}(\mathbf{r},\ t)\int\limits_{R\leqslant a}\frac{dV'}{R}=4\pi\ddot{\rho}\int\limits_0^a R\,dr=2\pi a^2\ddot{\rho},\tag{27}$$

а эта величина стремится к нулю при уменьшении a. Следовательно, из (24)—(26) вытекает, что при $a\to 0$

$$\nabla^2\varphi-\frac{1}{c^2}\ddot{\varphi}=\nabla^2(\varphi_1+\varphi_2)-\frac{1}{c^2}(\ddot{\varphi}_1+\ddot{\varphi}_2)\to-4\pi\rho(\mathbf{r},\ t),\tag{28}$$

так что (22) удовлетворяет неоднородному волновому уравнению для скалярного потенциала. Совершенно аналогичным способом можно показать, что каждая из декартовых компонент вектора (21) является решением соответствующего скалярного волнового уравнения, неоднородный член которого содержит вместо ρ соответствующую компоненту вектора $\mathbf{j}/c$. Следовательно, (21) удовлетворяет неоднородному волновому уравнению для векторного потенциала. Более того, в силу справедливости уравнения непрерывности (13) эти решения удовлетворяют также условию Лоренца (10).

Выражения (21) и (22) допускают простую физическую интерпретацию. Они показывают, что мы можем считать $\mathbf{A}$ и φ состоящими из вкладов от каждого элемента объема пространства, причем вклады от произвольного элемента

dV' в $\mathbf{A}$ и φ составляют соответственно

$$\frac{1}{c}\frac{\mathbf{j}(\mathbf{r}',\ t-R/c)}{R} \quad \text{и} \quad \frac{\rho(\mathbf{r}',\ t-R/c)}{R}.$$

Величина R/c в точности равна времени, необходимому для распространения света от точки $\mathbf{r}'$ до $\mathbf{r}$, так что каждый вклад должен «покинуть» элемент в такой предыдущий момент времени, чтобы «достичь» точки наблюдения в требуемое время t. По этой причине (21) и (22) называют *запаздывающими потенциалами* *).

Выражения (21) и (22) представляют собой частное решение волновых уравнений (11) и (12), а именно такое, которое получается при заданных зарядах и токах. Общее решение можно найти, если добавить к нему общее решение однородных волновых уравнений

$$\nabla^2\mathbf{A}-\frac{1}{c^2}\ddot{\mathbf{A}}=0, \tag{29}$$

$$\nabla^2\varphi-\frac{1}{c^2}\ddot{\varphi}=0, \tag{30}$$

также подчиняющееся условию Лорентца.

§ 2.2. Поляризация и намагничение

2.2.1. Выражение потенциалов через поляризацию и намагничение. В гл. 1 мы привлекали к изучению поля материальные уравнения $\mathbf{D}=\varepsilon\mathbf{E}$ и $\mathbf{B}=\mu\mathbf{H}$. Они означают, что поля $\mathbf{E}$ и $\mathbf{H}$ в каждой точке среды вызывают некие смещения $\mathbf{D}$ и $\mathbf{B}$, пропорциональные $\mathbf{E}$ и $\mathbf{H}$. Вместо того чтобы описывать взаимодействие поля и среды посредством таких «мультипликативных соотношений», мы будем описывать его с помощью «аддитивных соотношений»

$$\mathbf{D}=\mathbf{E}+4\pi\mathbf{P}, \tag{1}$$

$$\mathbf{B}=\mathbf{H}+4\pi\mathbf{M}. \tag{2}$$

Величина $\mathbf{P}$ называется *электрической поляризацией*, а $\mathbf{M}$ — *магнитной поляризацией* или *намагничением*. Физический смысл этих величин выяснится позже; здесь мы заметим только, что и $\mathbf{P}$ и $\mathbf{M}$ равны нулю в вакууме, и поэтому они отражают влияние среды на поле простым, интуитивно понятным способом.

Предположим, что среда является непроводящей ($\sigma=0$), и рассмотрим поле в области, где плотности тока и заряда равны нулю ($\rho=\mathbf{j}=0$). Исключая $\mathbf{D}$ и $\mathbf{H}$ из уравнений Максвелла (1.1.1)—(1.1.4) посредством (1) и (2), получим

$$\operatorname{rot}\mathbf{B}-\frac{1}{c}\dot{\mathbf{E}}=\frac{4\pi}{c}\tilde{\mathbf{j}}, \tag{3}$$

$$\operatorname{rot}\mathbf{E}+\frac{1}{c}\dot{\mathbf{B}}=0, \tag{4}$$

$$\operatorname{div}\mathbf{E}=4\pi\tilde{\rho}, \tag{5}$$

$$\operatorname{div}\mathbf{B}=0, \tag{6}$$

где *плотность свободного тока* $\tilde{\mathbf{j}}$ и *плотность свободного заряда* $\tilde{\rho}$ определяются соотношениями

$$\tilde{\mathbf{j}}=\dot{\mathbf{P}}+c\operatorname{rot}\mathbf{M}, \tag{7}$$

$$\tilde{\rho}=-\operatorname{div}\mathbf{P}. \tag{8}$$

---

*) Можно сконструировать решения также в форме *опережающих потенциалов* (с $(t+R/c)$ вместо $(t-R/c)$). Они представляют влияние приходящих сферических волн, тогда как запаздывающие потенциалы отражают влияние уходящих сферических волн.

За счет некоторой формальной асимметрии мы можем ограничиться решениями в виде одних запаздывающих потенциалов. При термодинамическом рассмотрении взаимодействия излучения с веществом оказывается, что асимметрия присуща физической ситуации. Это оправдывает наш выбор (см. [4], особенно стр. 26).

Уравнения (3)—(6) *формально идентичны уравнениям для поля в вакууме*, рассмотренным в предыдущем разделе. Следовательно, как и в случае вакуума, мы можем ввести такие векторный потенциал **A** и скалярный потенциал φ, что

$$\mathbf{B} = \operatorname{rot} \mathbf{A}, \tag{9}$$

$$\mathbf{E} = -\frac{1}{c}\dot{\mathbf{A}} - \operatorname{grad} \varphi. \tag{10}$$

Кроме того, если наложить, как и ранее, условие Лорентца

$$\operatorname{div} \mathbf{A} + \frac{1}{c}\dot{\varphi} = 0, \tag{11}$$

то по аналогии с (2.1.11) и (2.1.12) мы получим следующие уравнения для **A** и φ:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\ddot{\mathbf{A}} = -\frac{4\pi}{c}\tilde{\mathbf{j}}, \tag{12}$$

$$\nabla^2 \varphi - \frac{1}{c^2}\ddot{\varphi} = -4\pi\tilde{\rho}. \tag{13}$$

Условие (11) согласуется с (12) и (13), если

$$\frac{\partial\tilde{\rho}}{\partial t} + \operatorname{div}\tilde{\mathbf{j}} = 0. \tag{14}$$

Это соотношение удовлетворяется тождественно, что легко получить из (7) и (8), использовав векторное тождество $\operatorname{div} \operatorname{rot} \equiv 0$.

Согласно (2.1.21) и (2.1.22), а также (7) и (8) решения уравнений (12) и (13) можно выразить через поляризацию и намагничение в виде

$$\varphi = -\int \frac{1}{R}[\operatorname{div}' \mathbf{P}]\, dV', \tag{15}$$

$$\mathbf{A} = \int \frac{1}{R}\left[\operatorname{rot}' \mathbf{M} + \frac{1}{c}\dot{\mathbf{P}}\right] dV'. \tag{16}$$

Здесь дифференциальные операторы div' и rot' берутся для координат (x', y', z') точки интегрирования $\mathbf{r}'$ в элементе объема dV', а квадратные скобки означают *запаздывающие значения*, т. е. внутри каждой скобки аргумент t заменяется на $t - R/c$.

Прямой расчет позволяет получить следующие тождества:

$$\operatorname{div}' [\mathbf{P}] = [\operatorname{div}' \mathbf{P}] + \frac{1}{cR}\mathbf{R}\cdot[\dot{\mathbf{P}}], \tag{17а}$$

$$\operatorname{rot}' [\mathbf{M}] = [\operatorname{rot}' \mathbf{M}] + \frac{1}{cR}\mathbf{R}\times[\dot{\mathbf{M}}], \tag{17б}$$

$$\mathbf{R} = \mathbf{r} - \mathbf{r}'. \tag{18}$$

Следовательно, (15) и (16) можно записать в виде

$$\varphi = -\int\left\{\frac{1}{R}\operatorname{div}' [\mathbf{P}] - \frac{1}{cR^2}\mathbf{R}\cdot[\dot{\mathbf{P}}]\right\} dV', \tag{19}$$

$$\mathbf{A} = \int\left\{\frac{1}{R}\operatorname{rot}' [\mathbf{M}] - \frac{1}{cR^2}\mathbf{R}\times[\dot{\mathbf{M}}] + \frac{1}{cR}[\dot{\mathbf{P}}]\right\} dV'. \tag{20}$$

Из первого члена каждого интеграла можно выделить интеграл по граничной поверхности. Для этого воспользуемся векторными тождествами

$$\operatorname{div}\left(\frac{1}{R}\mathbf{P}\right) = \frac{1}{R}\operatorname{div}\mathbf{P} + \mathbf{P}\cdot\operatorname{grad}\frac{1}{R}, \tag{21а}$$

$$\operatorname{rot}\left(\frac{1}{R}\mathbf{M}\right) = \frac{1}{R}\operatorname{rot}\mathbf{M} - \mathbf{M}\times\operatorname{grad}\frac{1}{R}. \tag{21б}$$

Интегрируя приведенные соотношения по произвольной конечной области

и используя теорему Гаусса, получим

$$\int \frac{1}{R}(\mathbf{n}\cdot\mathbf{P})\,dS = \int\left\{\frac{1}{R}\operatorname{div}\mathbf{P} + \mathbf{P}\cdot\operatorname{grad}\frac{1}{R}\right\}dV, \tag{22а}$$

$$\int \frac{1}{R}(\mathbf{n}\times\mathbf{M})\,dS = \int\left\{\frac{1}{R}\operatorname{rot}\mathbf{M} - \mathbf{M}\times\operatorname{grad}\frac{1}{R}\right\}dV. \tag{22б}$$

Интегралы слева берутся по поверхности, ограничивающей эту область, $\mathbf{n}$ — единичный вектор внешней нормали к поверхности.

Предположим, что вещество (т. е. область пространства, где $\mathbf{P}$ и $\mathbf{M}$ отличны от нуля) находится внутри конечной замкнутой поверхности. Если интегралы (19) и (20) берутся по объему, заключенному внутри этой поверхности, то поверхностные интегралы в (22) равны нулю, а (19) и (20) можно записать в виде

$$\varphi = \int\left\{[\mathbf{P}]\cdot\operatorname{grad}'\frac{1}{R} + \frac{1}{cR^2}\mathbf{R}\cdot[\dot{\mathbf{P}}]\right\}dV', \tag{23}$$

$$\mathbf{A} = \int\left\{[\mathbf{M}]\times\operatorname{grad}'\frac{1}{R} - \frac{1}{cR^2}\mathbf{R}\times[\dot{\mathbf{M}}] + \frac{1}{cR}[\dot{\mathbf{P}}]\right\}dV'. \tag{24}$$

Последние выражения позволяют понять физический смысл величин $\mathbf{P}$ и $\mathbf{M}$. В самом деле, рассмотрим случай, когда $\mathbf{P}$ и $\mathbf{M}$ равны нулю всюду, кроме исчезающе малого элемента объема вблизи точки $\mathbf{r}_0(x_0, y_0, z_0)$. Формально это можно записать с помощью дельта-функции Дирака (см. приложение 4) в виде

$$\mathbf{P}(\mathbf{r}', t) = \mathbf{p}(t)\,\delta(\mathbf{r}'-\mathbf{r}_0), \tag{25}$$

$$\mathbf{M}(\mathbf{r}', t) = \mathbf{m}(t)\,\delta(\mathbf{r}'-\mathbf{r}_0). \tag{26}$$

Тогда (23) и (24) перейдут в

$$\varphi = [\mathbf{p}]\cdot\operatorname{grad}_0\frac{1}{R} + \frac{1}{cR^2}\mathbf{R}\cdot[\dot{\mathbf{p}}], \tag{27}$$

$$\mathbf{A} = [\mathbf{m}]\times\operatorname{grad}_0\frac{1}{R} - \frac{1}{cR^2}\mathbf{R}\times[\dot{\mathbf{m}}] + \frac{1}{cR}[\dot{\mathbf{p}}], \tag{28}$$

где

$$\mathbf{R} = \mathbf{r} - \mathbf{r}_0, \qquad R = |\mathbf{r} - \mathbf{r}_0|, \tag{29}$$

а grad_0 означает, что оператор берется относительно координат x_0, y_0, z_0 точки $\mathbf{r}_0$.

Уравнения (27) и (28) допускают простую интерпретацию. Рассмотрим электростатический потенциал двух фиксированных электрических зарядов $-e$ и $+e$, расположенных в точках, радиусы-векторы которых относительно фиксированной точки $Q_0(\mathbf{r}_0)$ равны $-\mathbf{a}/2$ и $\mathbf{a}/2$ соответственно. Кулоновский потенциал $\varphi(\mathbf{R})$, связанный с этими зарядами, равен (рис. 2.1)

$$\varphi = \frac{e}{R_2} - \frac{e}{R_1}. \tag{30}$$

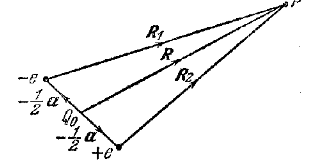


Рис. 2.1. К расчету потенциалов электрического диполя.

Если a достаточно мало, можно разложить $1/R_2$ по степеням компонент $\mathbf{a}$; имеем

$$\frac{1}{R_2} = \frac{1}{R_1} + \mathbf{a}\cdot\operatorname{grad}\frac{1}{R_1} + \ldots, \tag{31}$$

где операция grad берется по координатам точки, в которой расположен заряд $-e$. Пренебрегая в (31) членами более высоких порядков, получим вместо соотношения (30)

$$\varphi = e\mathbf{a}\cdot\operatorname{grad}\frac{1}{R_1} = e\frac{\mathbf{a}\cdot\mathbf{R}_1}{R_1^3}. \tag{32}$$

Пусть $\mathbf{a}$ постепенно уменьшается, а e неограниченно увеличивается таким

образом, что $e\mathbf{a}$ стремится к конечному значению $\mathbf{p}$, т. е.

$$\lim_{a \to 0} e\mathbf{a} = \mathbf{p}. \tag{33}$$

Тогда R_1 приближается к R, и в пределе (32) примет вид

$$\varphi = \mathbf{p} \cdot \operatorname{grad}_0 \frac{1}{R}. \tag{34}$$

В частном случае, когда $\mathbf{p}$ не зависит от времени, это выражение идентично (27).

Если заряды зависят от времени, но в каждый данный момент различаются лишь знаком, то вместо (30) получим

$$\varphi = \frac{e(t - R_2/c)}{R_2} - \frac{e(t - R_1/c)}{R_1}. \tag{35}$$

Снова переходя к пределу, найдем

$$\varphi = [\mathbf{p}] \cdot \operatorname{grad}_0 \frac{1}{R} + \frac{1}{cR^2} \mathbf{R} \cdot [\dot{\mathbf{p}}],$$

что совпадает с (27). Итак, (23) можно интерпретировать как скалярный потенциал *распределения электрических диполей с моментом* $\mathbf{P}$ *в единице объема.* Легко показать, что последний член в (28) является магнитным потенциалом, создаваемым этими диполями.

Аналогично можно показать, что (28) является векторным потенциалом магнитного диполя с моментом $\mathbf{m}(t)$, причем такой диполь, конечно, эквивалентен бесконечно малому замкнутому контуру с площадью A, нормальному к $\mathbf{m}$ и несущему ток *), равный $c\mathbf{m}/A$. Следовательно, два первых члена в (24) можно интерпретировать как векторный потенциал *распределения магнитных диполей с моментом* $\mathbf{M}$ *в единице объема.*

2.2.2. Векторы Герца. Вместо того чтобы использовать потенциалы $\mathbf{A}$ и φ, мы можем выразить поле через другую пару потенциальных функций $\mathbf{\Pi}_e$ и $\mathbf{\Pi}_m$, зависимость которых от $\mathbf{P}$ и $\mathbf{M}$ имеет значительно более простой вид. Они известны как *векторы Герца*, или *поляризационные потенциалы* **), и вводятся посредством соотношений

$$\mathbf{A} = \frac{1}{c} \dot{\mathbf{\Pi}}_e + \operatorname{rot} \mathbf{\Pi}_m, \tag{36}$$

$$\varphi = -\operatorname{div} \mathbf{\Pi}_e. \tag{37}$$

Как мы видим $\mathbf{\Pi}_e$ и $\mathbf{\Pi}_m$ связаны с $\mathbf{A}$ и φ такими же соотношениями, как поляризации $\mathbf{P}$ и $\mathbf{M}$ с $\tilde{j}/c$ и $\tilde{\rho}$.

Условие Лоренца при этом автоматически выполняется, а уравнения (12) и (13) для $\mathbf{A}$ и φ также будут удовлетворяться, если $\mathbf{\Pi}_e$ и $\mathbf{\Pi}_m$ являются решениями неоднородных волновых уравнений

$$\nabla^2 \mathbf{\Pi}_e - \frac{1}{c^2} \ddot{\mathbf{\Pi}}_e = -4\pi \mathbf{P}, \tag{38}$$

$$\nabla^2 \mathbf{\Pi}_m - \frac{1}{c^2} \ddot{\mathbf{\Pi}}_m = -4\pi \mathbf{M}. \tag{39}$$

Согласно § 2.1 частные их решения можно выразить через запаздывающие

---

*) Бесконечно малые токи Ампера рассматриваются в любом учебнике по электричеству, а также в книге [5] на стр. 305—306 (в русском переводе на стр. 411—412).

**) Векторы $\mathbf{\Pi}_e$ и $\mathbf{\Pi}_m$ преобразуются аналогично векторам $\mathbf{E}$ и $\mathbf{B}$ (или $\mathbf{D}$ и $\mathbf{H}$), т. е. образуют *6-вектор* (антисимметричный тензор второго ранга). Они являются обобщениями некой потенциальной функции, введенной Г. Герцем для электромагнитного поля осциллирующего диполя [6]. Векторный характер потенциала Герца был замечен А. Ричи [7], который ввел соответствующий магнитный потенциал. Общая теория векторов Герца и соответствующего калибровочного преобразования развита в работах [8] (см. также [9]).

потенциалы в виде

$$\mathbf{\Pi}_e=\int\frac{[\mathbf{P}]}{R}\,dV', \tag{40}$$

$$\mathbf{\Pi}_m=\int\frac{[\mathbf{M}]}{R}\,dV'. \tag{41}$$

Векторы поля можно выразить непосредственно через $\mathbf{\Pi}_e$ и $\mathbf{\Pi}_m$ простым дифференцированием. Из (9) и (36) получим

$$\mathbf{B}=\operatorname{rot}\left(\frac{1}{c}\dot{\mathbf{\Pi}}_e+\operatorname{rot}\mathbf{\Pi}_m\right), \tag{42}$$

а из (10), (36) и (37) найдем

$$\mathbf{E}=-\frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\dot{\mathbf{\Pi}}_e+\operatorname{rot}\mathbf{\Pi}_m\right)+\operatorname{grad}\operatorname{div}\mathbf{\Pi}_e=$$
$$=\operatorname{rot}\left(-\frac{1}{c}\dot{\mathbf{\Pi}}_m+\operatorname{rot}\mathbf{\Pi}_e\right)+\left(\nabla^2\mathbf{\Pi}_e-\frac{1}{c^2}\ddot{\mathbf{\Pi}}_e\right), \tag{43}$$

где использовано тождество grad div $\equiv$ rot rot $+\nabla^2$. Согласно (38) член во вторых скобках равен $-4\pi\mathbf{P}$. Следовательно, используя (1), находим

$$\mathbf{D}=\operatorname{rot}\left(-\frac{1}{c}\dot{\mathbf{\Pi}}_m+\operatorname{rot}\mathbf{\Pi}_e\right). \tag{44}$$

Наконец, из (2), (42) и (39) имеем

$$\mathbf{H}=\operatorname{rot}\left(\frac{1}{c}\dot{\mathbf{\Pi}}_e+\operatorname{rot}\mathbf{\Pi}_m\right)+\left(\nabla^2\mathbf{\Pi}_m-\frac{1}{c^2}\ddot{\mathbf{\Pi}}_m\right)=$$
$$=\frac{1}{c}\frac{\partial}{\partial t}\left(-\frac{1}{c}\dot{\mathbf{\Pi}}_m+\operatorname{rot}\mathbf{\Pi}_e\right)+\operatorname{grad}\operatorname{div}\mathbf{\Pi}_m. \tag{45}$$

В § 2.1 было показано, что потенциалы $\mathbf{A}$ и φ, связанные с данным полем, неоднозначны; любая другая пара ($\mathbf{A}'$, φ'), которую можно получить из возможной пары ($\mathbf{A}$, φ) посредством калибровочного преобразования (2.1.14), будет представлять то же поле. Векторы Герца также неоднозначны. Их можно подвергнуть следующему преобразованию, которое оставит векторы поля неизменными:

$$\mathbf{\Pi}'_e=\mathbf{\Pi}_e+\operatorname{rot}\mathbf{F}-\operatorname{grad}G,\qquad \mathbf{\Pi}'_m=\mathbf{\Pi}_m-\frac{1}{c}\dot{\mathbf{F}}, \tag{46}$$

где векторная функция $\mathbf{F}$ и скалярная функция G — произвольные решения однородных волновых уравнений

$$\nabla^2\mathbf{F}-\frac{1}{c^2}\ddot{\mathbf{F}}=0,\qquad \nabla^2G-\frac{1}{c^2}\ddot{G}=0. \tag{47}$$

Инвариантность векторов поля относительно такого преобразования легко показать, подставляя (46) в (36) и (37). Величины $\mathbf{A}$ и φ при этом преобразуются в $\mathbf{A}'$ и φ' согласно (2.1.14) с

$$\chi=-\frac{1}{c}\dot{G}.$$

Кроме того, подстановка (46) в (38) и (39) показывает, что $\mathbf{\Pi}'_e$ и $\mathbf{\Pi}'_m$ также удовлетворяют неоднородным волновым уравнениям для векторов Герца.

2.2.3. Поле линейного электрического диполя. Для дальнейшего полезно записать в явном виде полные решения уравнений поля в вакууме для поля, создаваемого линейным электрическим диполем, расположенным в точке $\mathbf{r}_0$ и колеблющимся в фиксированном направлении, определяемом единичным вектором $\mathbf{n}$. Такой диполь характеризуется электрической поляризацией

$$\mathbf{P}(\mathbf{r},\ t)=p(t)\,\delta(\mathbf{r}-\mathbf{r}_0)\,\mathbf{n}, \tag{48}$$

где p — функция времени, а δ — дельта-функция Дирака. В общем случае

диполь с ориентацией $\mathbf{n}$, также зависящей от времени, эквивалентен трем линейным диполям, дипольные моменты которых ориентированы вдоль трех взаимно перпендикулярных направлений.

Согласно (40) электрический вектор Герца, связанный с (48), равен

$$\mathbf{\Pi}_e = \frac{p(t - R/c)}{R}\mathbf{n}, \tag{49}$$

где R — расстояние точки $\mathbf{r}$ от точки $\mathbf{r}_0$.

Так как $\mathbf{\Pi}_m = 0$, а $\mathbf{\Pi}_e$ удовлетворяет всюду (кроме начала координат) однородному волновому уравнению для вакуума, уравнения (42)—(45) сводятся к

$$\mathbf{E} = \mathbf{D} = \operatorname{rot}\operatorname{rot}\mathbf{\Pi}_e, \tag{50}$$

$$\mathbf{B} = \mathbf{H} = \frac{1}{c}\operatorname{rot}\dot{\mathbf{\Pi}}_e. \tag{51}$$

Используя тождество $\operatorname{rot}\operatorname{rot} \equiv \operatorname{grad}\operatorname{div} - \nabla^2$ и соотношение $\nabla^2\mathbf{\Pi}_e = \frac{1}{c^2}\ddot{\mathbf{\Pi}}_e$, можно переписать (50) в виде

$$\mathbf{E} = \mathbf{D} = \operatorname{grad}\operatorname{div}\mathbf{\Pi}_e - \frac{1}{c^2}\ddot{\mathbf{\Pi}}_e. \tag{52}$$

Из (49) после простого расчета получим

$$\operatorname{div}\mathbf{\Pi}_e = -\left\{\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2}\right\}(\mathbf{n}\cdot\mathbf{R}),$$

$$\operatorname{grad}\operatorname{div}\mathbf{\Pi}_e = \left\{\frac{3[p]}{R^5} + \frac{3[\dot{p}]}{cR^4} + \frac{[\ddot{p}]}{c^2R^3}\right\}(\mathbf{n}\cdot\mathbf{R})\mathbf{R} - \left\{\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2}\right\}\mathbf{n},$$

$$\operatorname{rot}\mathbf{\Pi}_e = \left\{\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2}\right\}(\mathbf{n}\times\mathbf{R}),$$

где, как и раньше, величины в квадратных скобках соответствуют запаздывающим значениям. Подставляя эти выражения в (52) и (51), получим искомые формулы для векторов поля, а именно

$$\mathbf{E} = \mathbf{D} = \left\{\frac{3[p]}{R^5} + \frac{3[\dot{p}]}{cR^4} + \frac{[\ddot{p}]}{c^2R^3}\right\}(\mathbf{n}\cdot\mathbf{R})\mathbf{R} - \left\{\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2} + \frac{[\ddot{p}]}{c^2R}\right\}\mathbf{n}, \tag{53}$$

$$\mathbf{B} = \mathbf{H} = \left\{\frac{[\dot{p}]}{cR^3} + \frac{[\ddot{p}]}{c^2R^2}\right\}(\mathbf{n}\times\mathbf{R}). \tag{54}$$

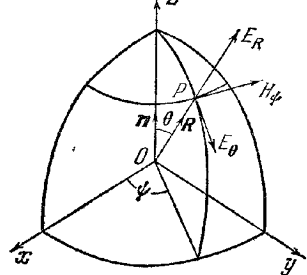


Рис. 2.2. Расчет поля линейного электрического диполя с дипольным моментом вдоль оси z.

Позже нам понадобятся также выражения для $\mathbf{E}$ и $\mathbf{H}$ в сферических координатах. Выбирая направление $\mathbf{n}$ в качестве оси z (рис. 2.2) и обозначая через $\mathbf{i}_R$, $\mathbf{i}_\theta$ и $\mathbf{i}_\psi$ единичные векторы в направлении увеличения R, θ и ψ, получим

$$\mathbf{R} = R\mathbf{i}_R, \quad \mathbf{n} = (\cos\theta)\,\mathbf{i}_R - (\sin\theta)\,\mathbf{i}_\theta, \quad \mathbf{n}\cdot\mathbf{R} = R\cos\theta, \quad \mathbf{n}\times\mathbf{R} = (R\sin\theta)\,\mathbf{i}_\psi. \tag{55}$$

Тогда (53) и (54) дают

$$\mathbf{E} = E_R\mathbf{i}_R + E_\theta\mathbf{i}_\theta, \qquad \mathbf{H} = H_\psi\mathbf{i}_\psi, \tag{56}$$

где

$$\left.\begin{aligned} E_R &= 2\left(\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2}\right)\cos\theta, \quad E_\theta = \left(\frac{[p]}{R^3} + \frac{[\dot{p}]}{cR^2} + \frac{[\ddot{p}]}{c^2R}\right)\sin\theta, \\ H_\psi &= \left(\frac{[\dot{p}]}{cR^2} + \frac{[\ddot{p}]}{c^2R}\right)\sin\theta. \end{aligned}\right\} \tag{57}$$

Таким образом, электрический вектор лежит в меридиональной плоскости, проходящей через ось диполя, а магнитный вектор перпендикулярен этой плоскости.

Особый интерес представляет поле в области, которая настолько удалена от диполя, что в приведенных выше уравнениях можно пренебречь всеми членами, кроме членов первого порядка относительно $1/R$. Эта *волновая* (или *радиационная*) *зона* характеризуется тем, что для нее

$$R \gg c\left|\frac{p}{\dot{p}}\right|, \qquad R \gg c\left|\frac{\dot{p}}{\ddot{p}}\right|. \tag{58}$$

В этой области

$$E_\theta \sim H_\psi \sim \frac{[\ddot{p}]}{c^2 R}\sin\theta, \tag{59}$$

а другие компоненты пренебрежимо малы. Следовательно, в волновой зоне векторы **E** и **H** по величине равны между собой и перпендикулярны друг к другу и к радиусу-вектору **R**, который совпадает там с направлением вектора Пойнтинга. Поэтому в этой зоне структура поля линейного электрического диполя подобна структуре поля плоской волны. Однако в случае диполя на каждой «волновой поверхности» (сфере с центром в точке $\mathbf{r}_0$) векторы поля меняются от точки к точке, уменьшаясь по величине от экватора к полюсам, причем на оси осциллятора они равны нулю. Поэтому диполь не излучает энергии в направлении своей оси.

Рассчитаем количество энергии, излучаемой за 1 *сек* через каждую сферическую волновую поверхность. Оно равно интегралу от величины вектора Пойнтинга S, взятому по этой поверхности. В волновой зоне

$$S = \frac{c}{4\pi}|\mathbf{E}\times\mathbf{H}| = \frac{c}{4\pi}|E_\theta H_\psi| = \frac{[\ddot{p}]^2}{4\pi c^3 R^2}\sin^2\theta. \tag{60}$$

Следовательно, общее количество энергии, протекающее через сферическую поверхность в секунду, равно ($d\sigma$ обозначает элемент поверхности)

$$\int S\,d\sigma = \frac{[\ddot{p}]^2}{4\pi c^3 R^2}\int_0^\pi \sin^2\theta\cdot 2\pi R^2\sin\theta\,d\theta = \frac{2[\ddot{p}]^2}{3c^3}. \tag{61}$$

Рассмотрим специальный случай, когда $p(t)$ является периодической функцией t с угловой частотой ω, т. е.

$$p(t) = p_0 e^{-i\omega t}, \tag{62}$$

где p_0 — комплексная постоянная. Будем считать, что вещественная часть (62) представляет p. Два условия (58) теперь сводятся к одному

$$R \gg \lambda/2\pi \qquad (\lambda = 2\pi c/\omega), \tag{63}$$

а (59) дает для неисчезающих компонент поля в волновой зоне выражение

$$E_\theta = H_\psi \sim -\left(\frac{\omega}{c}\right)^2 p_0\sin\theta\,\frac{\exp\{-i\omega(t-R/c)\}}{R}; \tag{64}$$

предполагается, что берется вещественная часть выражения, стоящего в правой части (64). Количество энергии, проходящее в секунду через единичную площадь сферической поверхности в волновой зоне, равно

$$S = \frac{c}{4\pi}\left(\frac{\omega}{c}\right)^4\frac{|p_0|^2}{R^2}\sin^2\theta\cos^2[\omega(t-R/c)-\alpha], \tag{65}$$

где α — фаза p_0.

Отсюда, усредняя по времени за интервал времени, большой по сравнению с периодом $T = 2\pi/\omega$, получим

$$\langle S\rangle = \frac{|p_0|^2}{8\pi R^2}\frac{\omega^4}{c^3}\sin^2\theta. \tag{66}$$

Поэтому энергия (усредненная по времени), проходящая в секунду через всю поверхность, равна

$$\int \langle S\rangle\, d\sigma = \frac{1}{3}\frac{|p_0|^2\,\omega^4}{c^3}. \tag{67}$$

§ 2.3. Формула Лорентц — Лоренца и элементарная теория дисперсии*)

2.3.1. Диэлектрическая и магнитная восприимчивости. Интерпретация потенциалов при помощи поляризации и намагничения имеет фундаментальное значение в теории атомного строения вещества. В рамках такой теории аддитивные соотношения (2.2.1) и (2.2.2), связывающие поле **E** с **D** и **P** и поле **H** с **B** и **M**, имеют более ясный физический смысл, чем мультипликативные соотношения (материальные уравнения (1.1.10) и (1.1.11)).

В этой теории вещество рассматривается как совокупность взаимодействующих частиц (атомов и молекул), находящихся в вакууме. Такие частицы образуют поле, которое испытывает большие локальные колебания внутри вещества. Это внутреннее поле видоизменяется любым полем, которое прикладывается извне, и свойства вещества находят путем усреднения по полному полю внутри него. Поскольку область, по которой проводится усреднение, велика по сравнению с линейными размерами частиц, их электромагнитные свойства можно отождествить со свойствами электрического и магнитного диполей; тогда вторичное поле совпадает с полем таких диполей (с запаздыванием). Фактически именно это мы только что описали, рассматривая вещество как некое непрерывное распределение, взаимодействующее с полем: такой подход соответствует первому приближению теории атомного строения (для медленного изменения в пространстве). В этом приближении для достаточно слабых полей **) мы можем предположить, что **P** и **M** пропорциональны соответственно **E** и **H**, т. е.

$$\mathbf{P} = \eta\mathbf{E}, \qquad \mathbf{M} = \chi\mathbf{H}. \tag{1}$$

Множитель η называют *диэлектрической восприимчивостью*, а χ — *магнитной восприимчивостью*, причем последняя является понятием, которое широко применяется в практических приложениях магнетизма. Сравнивая (1), (2.2.1) и (2.2.2) с материальными уравнениями (1.1.10) и (1.1.11), получим, что диэлектрическая проницаемость ε и магнитная проницаемость μ связаны с диэлектрической и магнитной восприимчивостями соотношениями

$$\varepsilon = 1 + 4\pi\eta, \qquad \mu = 1 + 4\pi\chi. \tag{2}$$

Посредством (2) обеспечивается полная формальная эквивалентность «мультипликативного» и «аддитивного» подходов. Но при таком формальном сравнении постоянных не используются никакие следствия из факта атомного строения вещества. Детальное рассмотрение, учитывающее результаты теории атомного строения, выходит за рамки настоящей книги, однако будет полезно ввести ряд необходимых понятий и формул. С их помощью мы достигнем более ясного понимания физического содержания интегральных уравнений, которые будут введены в § 2.4 вместо обычных дифференциальных уравнений теории Максвелла.

*) В настоящем разделе мы отклонимся от нашей дедуктивной схемы, чтобы достичь более ясного понимания явлений перед тем, как сформулировать интегральные уравнения, которые будут введены в § 2.4 вместо дифференциальных уравнений Максвелла. Рассмотрение здесь носит скорее иллюстративный, чем строгий характер, так как строгая теория включала бы большую часть теории атомного строения материи и увела бы нас далеко за разумные границы этой книги.

**) Поле считается сильным на атомном уровне, если электростатическая энергия электрона меняется на заметную долю ионизационного потенциала атома на расстоянии, равном диаметру атома, т. е. к сильным относят поля в миллионы вольт на сантиметр. В действительности же в реальных веществах электрический пробой наступает при значительно меньших полях.

Простейшее предположение, служащее первым шагом учета атомной структуры вещества, заключается в рассмотрении вещества как совокупности определенных физических объектов — молекул, которые могут поляризоваться и, следовательно, приобретать под действием внешнего поля электрический и магнитный моменты. В первом приближении можно предположить, что компоненты этих моментов являются линейными функциями компонент поля; в общем случае направление вектора момента не совпадает с направлением поля. Из этого предположения вытекает много следствий, на которых здесь мы остановимся лишь очень кратко. Мы ограничимся изотропными немагнитными веществами и рассмотрим вначале зависимость электрических постоянных от плотности для вещества, состоящего из одинаковых молекул. Мы исследуем также зависимость показателя преломления от частоты. Здесь мы приведем лишь несколько упрощенные рассуждения, а строгий, хотя и более формальный, вывод главного результата (формулы Лорентц — Лоренца) будет изложен в § 2.4.

2.3.2. Эффективное поле. Прежде всего необходимо различать *эффективное поле* **E**′, **H**′, действующее на молекулу, и *среднее*, или *наблюдаемое*, поле **E**, **H**, полученное усреднением по области, которая содержит множество молекул. Различие между этими двумя полями обусловлено промежутками между молекулами и зависит от числа молекул в единице объема (их концентрации N).

Для того чтобы оценить разность **E**′— **E**, рассмотрим отдельную молекулу и вообразим, что она окружена небольшой сферой, радиус которой, тем не менее, велик по сравнению с ее линейными размерами. Мы отдельно рассмотрим влияние на центральную молекулу вещества, находящегося снаружи и внутри этой сферы.

При определении влияния среды, расположенной *вне* сферы, очевидно, можно пренебречь молекулярной структурой и считать среду непрерывной. Тогда мы имеем право предположить, что вне сферы поляризация **P**, создаваемая средним электрическим полем, постоянна. Мы предположим далее, что молекулы, находящиеся *внутри* сферы, не создают какого-либо результирующего поля вблизи центральной молекулы. Это можно показать для ряда важных специальных случаев, в том числе для случая хаотического распределения. Следовательно, мы имеем право считать, что молекула расположена в такой сферической области, внутри которой вакуум, а вне — равномерно поляризованная среда. Теперь нужно определить потенциал φ подобной конфигурации, т. е. потенциал, создаваемый свободными зарядами, находящимися на сферической поверхности разрыва, на которой величина **P** меняется от значения, равного нулю внутри сферы, до постоянного значения вне сферы.

Для этой цели рассмотрим потенциал $\tilde{\varphi}$ «дополнительной» конфигурации, а именно однородно поляризованной сферы, находящейся в вакууме. Суперпозиция таких двух конфигураций дает равномерно поляризованную среду, свободную от всяких границ. Следовательно, потенциал, обусловленный границей, равен нулю, и мы имеем

$$\varphi+\tilde{\varphi}=0. \tag{3}$$

Потенциал $\tilde{\varphi}$ можно сразу получить из (2.2.23), считая **P** постоянным,

$$\varphi=-\tilde{\varphi}=-\mathbf{P}\cdot\int \operatorname{grad}' \frac{1}{R}\, dV'. \tag{4}$$

Так как

$$R=\sqrt{(x-x')^2+(y-y')^2+(z-z')^2},$$

мы можем в (4) заменить grad′ на —grad и написать

$$\varphi=-\tilde{\varphi}=\mathbf{P}\cdot\operatorname{grad}\int\frac{dV'}{R}=-\mathbf{P}\cdot\operatorname{grad}\varphi_0, \tag{5}$$

где

$$\varphi_0 = -\int \frac{dV'}{R}. \tag{6}$$

Последнее выражение можно рассматривать как потенциал однородно заряженной сферы с плотностью заряда, равной —1. Следовательно, оно удовлетворяет уравнению Пуассона

$$\nabla^2\varphi_0 = 4\pi. \tag{7}$$

Теперь компоненты поля, связанного с потенциалом (5), запишутся в виде

$$-\frac{\partial\varphi}{\partial x} = \frac{\partial}{\partial x}\left[P_x\frac{\partial\varphi_0}{\partial x} + P_y\frac{\partial\varphi_0}{\partial y} + P_z\frac{\partial\varphi_0}{\partial z}\right] = P_x\frac{\partial^2\varphi_0}{\partial x^2} + P_y\frac{\partial^2\varphi_0}{\partial x\,\partial y} + P_z\frac{\partial^2\varphi_0}{\partial x\,\partial z} \tag{8}$$

и т. д. Из условия симметрии в центре поля имеем

$$\frac{\partial^2\varphi_0}{\partial x\,\partial y} = \frac{\partial^2\varphi_0}{\partial y\,\partial z} = \frac{\partial^2\varphi_0}{\partial z\,\partial x} = 0 \tag{9}$$

и

$$\frac{\partial^2\varphi_0}{\partial x^2} = \frac{\partial^2\varphi_0}{\partial y^2} = \frac{\partial^2\varphi_0}{\partial z^2}. \tag{10}$$

Используя (7), находим, что каждый член равенств (10) равен $4\pi/3$; тогда выражение (8) показывает, что искомый вклад в эффективное поле равен

$$-\nabla\varphi = \frac{4\pi}{3}\mathbf{P}. \tag{11}$$

Полное поле внутри сферы, которое представляет собой эффективное поле, действующее на центральную молекулу, получается при добавлении к этой величине среднего поля $\mathbf{E}$, т. е.

$$\mathbf{E}' = \mathbf{E} + \frac{4\pi}{3}\mathbf{P}. \tag{12}$$

Соответствующее соотношение между $\mathbf{H}'$, $\mathbf{H}$ и $\mathbf{M}$ существует и для магнитных веществ. Однако, поскольку нас интересуют лишь немагнитные вещества, мы всегда будем полагать $\mathbf{H}' = \mathbf{H}$.

Далее свяжем поляризацию $\mathbf{P}$ с плотностью.

2.3.3. Средняя поляризуемость. Формула Лоренц — Лоренца. Как уже говорилось выше, мы будем предполагать, что для каждой молекулы электрический дипольный момент $\mathbf{p}$, возникающий под влиянием поля, пропорционален эффективному полю *) $\mathbf{E}'$, т. е.

$$\mathbf{p} = \alpha\mathbf{E}'. \tag{13}$$

Здесь предполагается, что рассматриваемая нами молекула изотропна **), но поскольку нас интересует лишь эффект, усредненный по всем возможным ее ориентациям, нет необходимости считать каждую *отдельную* молекулу изотропной. В этом случае α будет представлять *среднюю поляризуемость*, а так как $\mathbf{p}$ имеет размерность $[el]$, $\mathbf{E}'$ — размерность $[el^{-2}]$ (e — заряд, l — длина), то α будет иметь размерность $[l^3]$, т. е. размерность объема.

Если, как и раньше, N — число молекул в единице объема, то полный *электрический момент* $\mathbf{P}$ *единицы объема* будет равен

$$\mathbf{P} = N\mathbf{p} = N\alpha\mathbf{E}'. \tag{14}$$

Исключая $\mathbf{E}'$ из уравнений (12) и (14), получим первую из формул (1), причем

*) Мы ограничимся здесь рассмотрением так называемых неполярных молекул, не обладающих постоянным дипольным моментом в отсутствие поля.

**) В общем случае поляризуемость α является тензором второго ранга.

выражение для диэлектрической восприимчивости запишется в явном виде следующим образом:

$$\eta = \frac{N\alpha}{1 - \frac{4\pi}{3} N\alpha}. \tag{15}$$

Подставляя это значение η в первую из формул (2), находим для диэлектрической проницаемости

$$\varepsilon = \frac{1 + \frac{8\pi}{3} N\alpha}{1 - \frac{4\pi}{3} N\alpha}. \tag{16}$$

В свою очередь (16) дает зависимость средней поляризуемости от ε и N или (если использовать соотношение Максвелла $\varepsilon = n^2$) от n и N, а именно:

$$\alpha = \frac{3}{4\pi N} \frac{\varepsilon - 1}{\varepsilon + 2} = \frac{3}{4\pi N} \frac{n^2 - 1}{n^2 + 2}. \tag{17}$$

По удивительному совпадению соотношение (17) было получено независимо и практически одновременно двумя учеными с почти идентичными фамилиями, Лорентцом [10] и Лоренцом [11], и поэтому называется *формулой Лорентц — Лоренца* *). Как мы видим, она служит мостом, связывающим феноменологическую теорию Максвелла с теорией атомного строения вещества.

Вместо средней поляризуемости α часто пользуются другой величиной A, называемой *молекулярной рефракцией* (в случае одноатомных веществ она называется также *атомной рефракцией*). Это, по существу, полная поляризуемость моля вещества **), определяемая как

$$A = \frac{4\pi}{3} N_m \alpha. \tag{18}$$

Здесь $N_m = 6{,}02 \cdot 10^{23}$— число Авогадро, равное числу молекул в одном моле. Если W — молекулярный вес, ρ — плотность, p — давление и T — абсолютная температура, то молярный объем равен

$$\frac{N_m}{N} = \frac{W}{\rho} = \frac{RT}{p}, \tag{19}$$

где использован закон Бойля. Следовательно, молекулярную рефракцию можно записать в двух формах

$$A = \frac{W}{\rho} \frac{n^2 - 1}{n^2 + 2} = \frac{RT}{p} \frac{n^2 - 1}{n^2 + 2}. \tag{20}$$

Выражение (20) показывает, что A имеет размерность и порядок величины молярного объема.

Для газа n^2 почти равно единице, так что

$$\alpha \approx \frac{n^2 - 1}{4\pi N}, \qquad A \approx \frac{W}{\rho} \frac{n^2 - 1}{3} = \frac{RT}{p} \frac{n^2 - 1}{3}. \tag{21}$$

---

*) Аналогичная формула для статических полей была раньше получена в работах [12] и [13]. Их авторы пытались объяснить диэлектрические свойства изоляторов, предположив, что атомы являются небольшими проводящими сферами, взаимное расстояние между которыми велико по сравнению с их диаметрами; они получили следующее соотношение между концентрацией сфер g и диэлектрической проницаемостью:

$$g = \frac{\varepsilon - 1}{\varepsilon + 2}.$$

**) Молем или грамм-молекулой называется такое количество вещества, вес которого в граммах равен молекулярному весу этого вещества.

Формула (20) дает в явном виде зависимость показателя преломления от плотности для света какого-то определенного цвета. Она должна выполняться при изменении плотности при условии, что сохраняется изотропия. В случае *газов* мы действительно обнаруживаем, что n^2-1 и ρ почти пропорциональны друг другу, как это следует из второй формулы (21). И даже при высоком давлении, когда n заметно отличается от единицы, молекулярная рефракция остается по существу постоянной, что видно из табл. 2.1. Оказывается, что молекулярная рефракция остается также практически постоянной и при конденсации газа в жидкость (табл. 2.2).

Таблица 2.1

Молекулярная рефракция A воздуха при температуре около 14,5° С для различных давлений (D-линия натрия)

Давление *атм*	n	A
1,00	1,0002929	4,606
42,13	1,01241	4,661
96,16	1,02842	4,713
136,21	1,04027	4,743
176,27	1,05213	4,772

Используя (20), можно показать, что молекулярная рефракция смеси двух веществ в хорошем приближении равна сумме вкладов от каждого вещества. Таким образом, если смешиваются две жидкости с рефракциями A_1 и A_2 и если единичный объем первой жидкости содержит N_1 молекул, а второй — N_2 молекул, то молекулярная рефракция смеси будет равна

$$A=\frac{N_1A_1+N_2A_2}{N_1+N_2}. \tag{22}$$

Таблица 2.2

Молекулярная рефракция A различных соединений для D-линии натрия

Вещество	Формула	W	n	ρ	A
Кислород	O_2	32	1,221 1,000271	1,124	4,00 4,05
Соляная кислота	HCl	36,5	1,245 1,000447	0,95	5,95 6,68
Вода	H_2O	18	1,334 1,000249	1,00	3,71 3,72
Сероуглерод	CS_2	76	1,628 1,00147	1,264	21,33 21,99
Ацетон	C_3H_6O	58	1,3589 1,00108	0,791	16,14 16,16

В столбцах, в которых приведены величины n и A, первое значение относится к жидкости, второе к пару. Плотности паров неизвестны, но рассчитаны на основе закона идеальных газов из известной плотности водорода

Таблица 2.3

Молекулярная рефракция для D-линии натрия смеси воды и серной кислоты при 15° С

Относительное содержание H_2SO_4, вес %	$\frac{N_1}{N_1+N_2}$	W	n	ρ	A (получено по (22))	A (рассчитано по (20))
0	0	18,00	1,3336	0,9991	3,72	3,72
19,98	0,044	21,52	1,3578	1,1381	4,19	4,15
39,76	0,109	26,72	1,3817	1,2936	4,80	4,81
59,98	0,216	35,28	1,4065	1,4803	5,86	5,86
80,10	0,425	52,00	1,4308	1,6955	7,95	7,93
100	1	98	1,4277	1,8417	13,68	13,68

Этот результат также оказывается в достаточно хорошем согласии с экспериментом (табл. 2.3).

Наконец, рассмотрим зависимость молекулярной рефракции A соединения от атомных рефракций составляющих его атомов. Если молекула состоит из $\mathcal{N}_1$ атомов с рефракцией A_1, $\mathcal{N}_2$ атомов с рефракцией A_2 и т. д., то, очевидно,

$$A = \mathcal{N}_1 A_1 + \mathcal{N}_2 A_2. \qquad (23)$$

Эта формула также довольно хорошо согласуется с экспериментом, как видно из табл. 2.5. Рефракции составляющих атомов приведены в табл. 2.4.

Таблица 2.4

Атомные рефракции различных элементов для D-линии натрия

Элемент	Символ	Атомный вес	A
Водород	H	1	1,02
Углерод	C	12	2,11
Кислород	O	16	2,01
Сера	S	32	8,23
Хлор	Cl	35,5	5,72

Таблица 2.5

Молекулярные рефракции различных соединений для D-линии натрия

Вещество	Формула	A (рассчитано по (20))	A (рассчитано по (23))
Соляная кислота	HCL	6,68	6,74
Вода	H_2O	3,72	4,05
Сероуглерод	CS_2	21,97	18,57
Ацетон	C_3H_6O	16,14	14,46

Согласие, полученное в этом случае, лучше, чем можно было бы ожидать, исходя из нашего простого рассмотрения. В других случаях действительно нужно учитывать некоторые другие соображения, например, при исследовании азота мы должны приписать атому азота различные атомные поляризуемости в зависимости от того, в какие соединения он входит. Вводя необходимые модификации, мы видим, что α служит хорошо определенным атомным параметром, который можно использовать также при изучении других явлений, таких, как силы сцепления Ван-дер-Ваальса, поверхностная адсорбция и т. д. Эти явления, однако, удовлетворительно объясняются лишь с помощью квантовой механики.

2.3.4. Элементарная теория дисперсии. В § 1.2 указывалось, что фазовая скорость и, следовательно, показатель преломления не могут быть постоянными среды, как мы предполагали вначале при формальном рассмотрении. Эти величины должны зависеть от частоты. Изменение показателя преломления с частотой составляет явление *дисперсии*. Для адекватного описания дисперсии необходимо сильно углубиться в теорию атомного строения материи, но можно дать упрощенную модель диспергирующей среды, используя один или два основных результата, касающихся структуры молекул.

Молекула состоит из нескольких тяжелых частиц (ядер атомов, образующих молекулу), вокруг которых обращаются легкие частицы (электроны). Электроны несут отрицательный заряд, а ядра — положительный. В нейтральных молекулах заряды электронов точно компенсируют заряды ядер. Однако центры положительных (ядерных) и центры отрицательных (электронных) зарядов могут не совпадать; такая система является электрическим диполем и называется *полярной*. Для простоты мы здесь не будем рассматривать полярные молекулы, хотя они играют огромную роль во многих физических и химических явлениях.

Если неполярную молекулу поместить в электрическое поле, то электроны и ядра смещаются и возникает дипольный момент. Векторная сумма всех дипольных моментов молекул в единице объема является по существу вектором поляризации **P**, введенным формально в предыдущем разделе.

Для того чтобы определить зависимость поляризации и показателя преломления от частоты поля, нам прежде всего необходимо найти смещение $\mathbf{r}$ каждой заряженной частицы относительно ее положения равновесия. Мы можем предположить, что на каждый электрон действует сила Лорентца $\mathbf{F}$ (см. уравнение (1.1.34)), равная

$$\mathbf{F}=e\left(\mathbf{E}'+\frac{\mathbf{v}}{c}\times\mathbf{B}'\right),$$

где e — заряд электрона, а $\mathbf{v}$ — его скорость. Пусть скорость электрона мала по сравнению со скоростью света в вакууме c, и, следовательно, в выражении для силы Лорентца можно пренебречь вкладом магнитного поля. Строгое определение эффективного смещения ядер и электронов под действием электрической силы является сложной проблемой квантовой механики. Однако представляется правдоподобным (и действительно подтверждается строгой теорией), что с хорошим приближением электроны ведут себя так, как если бы при отклонении от положения равновесия на них действовала квазиупругая возвращающая сила $\mathbf{Q}=-q\mathbf{r}$. Следовательно, если через m обозначить массу электрона, то уравнение его движения запишется в виде

$$m\ddot{\mathbf{r}}+q\mathbf{r}=e\mathbf{E}'. \tag{24}$$

Пусть ω — циклическая частота поля падающей волны, т. е.

$$\mathbf{E}'=\mathbf{E}_0'\exp(-i\omega t); \tag{25}$$

будем искать решение (24) в виде

$$\mathbf{r}=\mathbf{r}_0\exp(-i\omega t). \tag{26}$$

Тогда искомое стационарное решение уравнения (24) запишется следующим образом:

$$\mathbf{r}=\frac{e\mathbf{E}'}{m(\omega_0^2-\omega^2)}, \tag{27}$$

где

$$\omega_0=\sqrt{\frac{q}{m}} \tag{28}$$

называется *резонансной частотой*. Согласно (27) электрон колеблется с частотой поля падающей волны.

Каждый электрон вносит в поляризацию момент $\mathbf{p}=e\mathbf{r}$. Ядра также вносят свой вклад в нее, но так как массы ядер велики по сравнению с массой электронов, их вкладом в первом приближении можно пренебречь. Предположим сначала, что в молекуле имеется лишь один эффективный электрон с резонансной частотой ω_0; тогда для полной поляризации $\mathbf{P}$ мы получим выражение

$$\mathbf{P}=N\mathbf{p}=Ne\mathbf{r}=N\frac{e^2}{m}\frac{\mathbf{E}'}{\omega_0^2-\omega^2}. \tag{29}$$

Сравнение (29) с (14) дает

$$N\alpha=N\frac{e^2}{m(\omega_0^2-\omega^2)}, \tag{30}$$

что выражает «плотность поляризуемости» через атомные параметры. Таким образом, оказывается, что величина α непостоянна, как это было бы, если бы ε в (17) означало статическую диэлектрическую проницаемость. Удобно ввести понятие зависящей от частоты диэлектрической проницаемости $\varepsilon(\omega)$, которая определяется соотношением Максвелла $\varepsilon=n^2$, где n — показатель преломления, являющийся функцией от ω, т. е. $n(\omega)$. Тогда статическая диэлектрическая проницаемость равна величине $\varepsilon(0)=n^2(0)$; согласно (17) она соответствует предельному значению $N\alpha(0)$, которое в соответствии с (30) имеет вид

$$N\alpha(0)=\frac{Ne^2}{m\omega_0^2}.$$

Подставив это значение в (16), мы получим статическую диэлектрическую проницаемость $\varepsilon(0)$.

Для $\omega \neq 0$ функция $N\alpha(\omega)$, согласно (30), монотонно увеличивается с ростом ω и обращается в бесконечность (резонансная точка) при $\omega = \omega_0$; для $\omega > \omega_0$ при увеличении ω она стремится к нулю с отрицательной стороны. Подставляя (30) в (17), мы найдем зависимость показателя преломления от частоты в явном виде

$$\frac{n^2-1}{n^2+2} = \frac{4\pi}{3}\frac{Ne^2}{m(\omega_0^2-\omega^2)}. \tag{31}$$

Для газа n близко к единице *), и поэтому в знаменателе левой части мы можем положить $n^2+2 \sim 3$; тогда

$$n^2-1 \approx 4\pi N\alpha = \frac{4\pi Ne^2}{m(\omega_0^2-\omega^2)}. \tag{32}$$

Как мы видим, n растет с увеличением частоты. Говорят, что в этом случае имеет место *нормальная* дисперсия. Далее, n больше или меньше единицы в зависимости от того, меньше ω, чем ω_0, или больше нее, и n приближается к единице при увеличении ω (рис. 2.3) **).

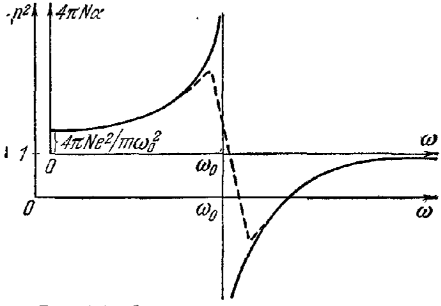


Рис. 2.3. Диперсионные кривые для газа.

Сплошная кривая — влияние затухания не учитывается; пунктирная кривая — с учетом затухания. В последнем случае по оси ординат отложены вещественные части величин n^2 и $4\pi N\alpha$.

На самом же деле при резонансной частоте ($\omega = \omega_0$) n и α не бесконечны, как утверждает наша формула. Возникает лишь формальная особенность, так как мы пренебрегали влиянием затухания. Фактически затухание служит существенным фактором во всем процессе, поскольку колеблющиеся электроны излучают электромагнитные волны, которые уносят энергию. Имеются и другие причины диссипации энергии (например, соударения между атомами). Формально затухание можно учесть путем добавления в уравнение движения (24) члена $g\dot{\mathbf{r}}$, представляющего тормозящую силу

$$m\ddot{\mathbf{r}} + g\dot{\mathbf{r}} + q\mathbf{r} = e\mathbf{E}'. \tag{33}$$

Тогда вместо (27) мы получим

$$\mathbf{r} = \frac{e\mathbf{E}'}{m(\omega_0^2-\omega^2) - i\omega g}. \tag{34}$$

Поляризация, а следовательно, и $N\alpha$ становятся комплексными. Можно показать (аналогичная ситуация возникает и в теории металлов, где используется комплексный показатель преломления, см. гл. 13), что вещественная часть этой комплексной функции является истинной «плотностью поляризуемости». Она показана пунктирной кривой на рис. 2.3. Кривая имеет резкий максимум при значении ω, несколько меньшем ω_0, и резкий минимум при значении, несколько большем ω_0. Между максимумом и минимумом функция умень-

*) Речь идет здесь о частотах, не слишком близких к частоте ω_0. При $\omega \to \omega_0$, как видно из (32), $n \to \infty$. (*Прим. перев.*)

**) Поскольку здесь идет речь о вещественном показателе преломления n, формула (32), в которой не учтено затухание, не может правильно передавать поведение n вблизи частоты ω_0. В самом деле, из (32) при $\omega \gtrsim \omega_0$ следует, что $n^2 < 0$ и, значит, величина n должна быть комплексной. (*Прим. перев.*)

шается с увеличением частоты, и мы говорим об области *аномальной дисперсии*. В этом случае лучи с более короткой длиной волны преломляются меньше, чем лучи с бóльшей длиной волны, что приводит к обращению обычного порядка чередования призматических цветов. Для наших целей область аномальной дисперсии не представляет большого интереса, так как резонансные частоты свободных атомов лежат почти исключительно в ультрафиолетовой области спектра. Показатель преломления в видимой области спектра в таком случае всегда больше единицы.

До сих пор мы предполагали, что у системы имеется лишь одна резонансная частота. В общем случае будет существовать много таких частот даже в системе, состоящей из молекул одного сорта; тогда (31) и (32) придется заменить более общими выражениями. Вновь пренебрегая временно движением ядер, мы получим вместо (31) соотношение

$$\frac{4\pi}{3}N\alpha=\frac{n^2-1}{n^2+2}=\frac{4\pi}{3}N\frac{e^2}{m}\sum_k\frac{f_k}{\omega_k^2-\omega^2}, \tag{35}$$

где Nf_k — число электронов для соответствующей резонансной частоты ω_k. Для газов ($n\sim 1$) последнее соотношение можно переписать в виде

$$n^2-1=4\pi N\alpha=\sum_k\frac{\rho_k}{\nu_k^2-\nu^2}=\sum_k\frac{\rho_k}{c^2}\frac{\lambda^2\lambda_k^2}{\lambda^2-\lambda_k^2}, \tag{36}$$

где

$$\rho_k=N\frac{e^2}{\pi m}f_k,\qquad \nu_k=\frac{\omega_k}{2\pi}=\frac{c}{\lambda_k},\qquad \nu=\frac{\omega}{2\pi}=\frac{c}{\lambda}. \tag{37}$$

Используя тождество

$$\frac{\lambda^2}{\lambda^2-\lambda_k^2}=1+\frac{\lambda_k^2}{\lambda^2-\lambda_k^2},$$

получим вместо (36)

$$n^2-1=a+\sum_k\frac{b_k}{\lambda^2-\lambda_k^2}, \tag{38}$$

где

$$a=\sum_k\frac{\rho_k}{\nu_k^2}=\frac{1}{c^2}\sum_k\rho_k\lambda_k^2,\qquad b_k=\frac{c\rho_k}{\nu_k^4}=\frac{1}{c^2}\rho_k\lambda_k^4. \tag{39}$$

Для представления показателя преломления во всей видимой области спектра обычно достаточно учесть лишь одну или две резонансные частоты в ультрафиолетовой области Кох [14], например, нашел, что в интервале от $\lambda=0{,}436$ *мк* до $\lambda=8{,}68$ *мк* для водорода, кислорода и воздуха выполняется следующее соотношение:

$$n^2-1=a+\frac{b}{\lambda^2-\lambda_0^2}; \tag{40}$$

постоянные a, b и λ_0 приведены в табл. 2.6.

Таблица 2.6

Дисперсионные постоянные для кислорода, водорода и воздуха в области 0,436 — 8,68 *мк* при 0°С и 760 *мм рт. ст.*

Газ	$a\cdot 10^5$	$b\cdot 10^{14}$, *см*^{2}	λ_0^2, 10^{-8} *см*^{2}	ν_0, 10^{15} *сек*$^{-1}$
Водород	27,216	211,2	0,007760	3,40
Кислород	52,842	369,9	0,007000	3,55
Воздух	57,642	327,7	0,005685	3,98

В спектральном интервале, который не содержит резонансных частот, выражение (38) с хорошей точностью можно заменить более простым. Для всех веществ, которые кажутся глазу прозрачными, таким интервалом является *видимая* область спектра. Обозначим через ν_v резонансные частоты с коротковолновой (фиолетовой) стороны нашего интервала, а через ν_r— частоты с длинноволновой (красной) его стороны; тогда дисперсионная формула (36) при разложении в степенные ряды относительно ν и λ соответственно примет вид

$$\left.\begin{aligned} n^2-1 &= A+B\nu^2+C\nu^4+\ldots-\frac{B'}{\nu^2}-\frac{C'}{\nu^4}-\ldots= \\ &= A+\frac{Bc^2}{\lambda^2}+\frac{Cc^4}{\lambda^4}+\ldots-\frac{B'\lambda^2}{c^2}-\frac{C'\lambda^4}{c^4}-\ldots, \end{aligned}\right\} \quad (41)$$

где

$$\left.\begin{aligned} &A=\sum_v \frac{\rho_v}{\nu_v^2}, \quad B=\sum_v \frac{\rho_v}{\nu_v^4}, \quad C=\sum_v \frac{\rho_v}{\nu_v^6}, \ \ldots \\ &B'=\sum_r \rho_r, \quad C'=\sum_r \rho_r\nu_r^2, \ \ldots \end{aligned}\right\} \quad (42)$$

В такой области, свободной от поглощения, значение n для газов так мало отличается от единицы, что мы можем заменить n^2-1 на $2(n-1)$. Кроме того, члены с B', C', ..., возникающие из-за резонансных частот, соответствующих ультрафиолетовой области, обычно не оказывают заметного влияния. Тогда, оставив только члены до $1/\lambda^2$, увидим, что (41) переходит в *формулу Коши* [15]

$$n-1=A_1\left(1+\frac{B_1}{\lambda^2}\right), \quad (43)$$

где

$$A_1=\frac{A}{2}, \quad B_1=\frac{Bc^2}{A}. \quad (44)$$

В табл. 2.7 приведены значения A_1 и B_1 для наиболее важных газов. Для иллюстрации точности формулы (43) в табл. 2.8 сравниваются друг с другом

Таблица 2.7

Значения постоянных A_1 и B_1 в дисперсионной формуле Коши для различных газов

Газ	$A_1\cdot10^5$	$B_1\cdot10^{9}$, $см^2$	Газ	$A_1\cdot10^5$	$B_1\cdot10^{9}$, $см^2$
Аргон	27,92	5,6	Кислород	26,63	5,07
Азот	29,19	7,7	Воздух	28,79	5,67
Гелий	3,48	2,3	Этан	73,65	9,08
Водород	13,6	7,7	Метан	42,6	14,41

Таблица 2.8

Наблюдаемые значения показателя преломления для воздуха и значения, даваемые дисперсионной формулой Коши (43)

$\lambda\cdot10^5$, см	$(n-1)\,10^4$		Разность	$\lambda\cdot10^5$, см	$(n-1)\,10^4$		Разность
	наблюдаемые	рассчитанные			наблюдаемые	рассчитанные	
7,594	2,905	2,907	0,002	4,308	2,966	2,967	0,001
6,563	2,916	2,917	0,001	3,969	2,983	2,983	0,000
5,896	2,926	2,926	0,000	3,728	2,995	2,996	0,001
5,378	2,935	2,935	0,000	3,441	3,016	3,017	0,001
5,184	2,940	2,940	0,000	3,180	3,040	3,041	0,001
4,861	2,948	2,948	0,000	3,021	3,056	3,058	0,002
4,677	2,951	2,954	0,003	2,948	3,065	3,067	0,002

наблюдаемые значения показателя преломления для воздуха и значения, даваемые формулой Коши.

В случае веществ с большой плотностью, т. е. для жидкостей или твердых тел, нельзя заменять n на единицу в знаменателе второго члена (35). Тем не менее можно привести (35) к той же форме, что и выше. Так как

$$\frac{n^2-1}{n^2+2}=\frac{4\pi}{3}N\alpha,$$

то

$$n^2-1=\frac{12\pi N\alpha}{3-4\pi N\alpha}, \tag{45}$$

где

$$4\pi N\alpha=\frac{Ne^2}{\pi m}\sum_k \frac{f_k}{\nu_k^2-\nu^2}=\sum_k \frac{\rho_k}{\nu_k^2-\nu^2}. \tag{46}$$

Как правило, достаточно учитывать лишь конечное число резонансных частот. Отсюда следует, что величина n^2-1, определяемая формулой (45), представляет собой рациональную функцию ν^2, и поэтому ее можно разложить на элементарные дроби. Для этой цели мы должны найти значения, при которых знаменатель обращается в нуль, т. е. корни уравнения

$$3-4\pi N\alpha=3-\sum_k \frac{\rho_k}{\nu_k^2-\nu^2}=0. \tag{47}$$

Обозначив их через $\bar{\nu}_k$, можно привести (45) к виду

$$n^2-1=\sum_k \frac{\bar{\rho}_k}{\bar{\nu}_k^2-\nu^2}. \tag{48}$$

Это выражение идентично по форме с формулой (36) для газов.

Например, при наличии лишь одной резонансной частоты $\bar{\nu}_1$ она будет корнем уравнения

$$3-\frac{\rho_1}{\nu_1^2-\nu^2}=0, \tag{49}$$

откуда

$$\bar{\nu}_1^2=\nu_1^2-\frac{1}{3}\rho_1, \qquad \bar{\rho}_1=\rho_1. \tag{50}$$

Уравнение (48), которое можно также переписать в форме (38), известно как *дисперсионная формула Зельмейера*.

До сих пор мы пренебрегали влиянием движения ядер. В действительности оно существенно лишь при очень длинных волнах (инфракрасная область). Причину этого легко понять. Электрический момент и среднюю поляризуемость приближенно можно разложить на две части — одну, связанную с электронами, и другую, связанную с ядрами, т. е.

$$\mathbf{p}=\mathbf{p}_e+\mathbf{p}_n, \tag{51}$$

где

$$\mathbf{p}_e=\alpha_e\mathbf{E}', \qquad \mathbf{p}_n=\alpha_n\mathbf{E}'. \tag{52}$$

Электроны почти мгновенно будут следовать за полем вплоть до довольно высоких частот, включая во многих случаях частоты всего видимого спектра. Масса же ядер, так велика, что при высоких частотах они не могут следовать за полем, т. е. для видимого света $\alpha_n\approx 0$. Это видно также из дисперсионной формулы (41), где n^2-1 представляется двумя группами членов — одной, связанной с высокочастотными («фиолетовыми») колебаниями ν_v, и другой — с низкочастотными («красными») колебаниями ν_r. Разумно предположить (и это строго следует из квантовой механики), что квазиупругие силы, связывающие ядра и электроны, одинаковы по порядку величины; следовательно, по порядку

величины частоты будут связаны соотношением

$$m\nu_e^2 \sim M\nu_n^2, \qquad (53)$$

где m — масса электрона, M — масса ядра, ν_e и ν_n— частоты колебаний электронов и ядер соответственно. Для того чтобы обобщить формулы (41) и (42) и учесть движение ядер, мы можем просто предположить, что «фиолетовые» колебания обусловлены электронами, а «красные» колебания — ядрами. Тогда придется также различать два типа коэффициентов ρ_k, определяемых выражением (37) — одни коэффициенты будут связаны с электронными, другие — с ядерными колебаниями. При этом по порядку величины выполняется соотношение

$$m\rho_e \sim M\rho_n. \qquad (54)$$

Следовательно, значения коэффициентов (42) дисперсионных формул будут равны по порядку величины

$$B \sim \frac{A}{\nu_e^2}, \quad C \sim \frac{A}{\nu_e^4}, \quad B' \sim \frac{m}{M} A\nu_e^2, \quad C' \sim \frac{m}{M} A\nu_e^4, \qquad (55)$$

откуда

$$n^2-1 \sim A\left\{1+a\frac{\nu^2}{\nu_e^2}+b\frac{\nu^4}{\nu_e^4}+\ldots-\frac{m}{M}\left(a'\frac{\nu_e^2}{\nu^2}+b'\frac{\nu_e^4}{\nu^4}+\ldots\right)\right\}, \qquad (56)$$

где a, b, ..., a', b', ...— численные постоянные порядка единицы. Первая группа членов соответствует электронной части поляризуемости α_e, вторая группа — ядерной части α_n. Мы видим, что, поскольку величина m/M мала ($\approx 1/1840$), члены второй группы пренебрежимо малы при условии, что частота ν составляет не очень малую долю «фиолетовых» резонансных частот ν_e.

Следовательно, для оптической частоты (т. е. частоты, соответствующей видимому свету) поляризация выражается, по существу, только через α_e (как предполагалось в наших предыдущих расчетах), тогда как для статического поля она выражается через $\alpha_e+\alpha_n$.

Рассмотрение в настоящем разделе основывалось исключительно на классической механике. Когда подобные расчеты выполняются на основе квантовой механики, взаимодействие поля с веществом по-прежнему можно описывать при помощи виртуальных осцилляторов, но число их, даже при наличии лишь одного электрона, оказывается бесконечно большим. И хотя уравнение (35) все еще остается применимым, но силы осцилляторов f_k определяют теперь не число электронов некоторого типа, а скорее число виртуальных осцилляторов, принадлежащих одному электрону или группе электронов. В большинстве случаев заметную величину имеет лишь конечное число значений f_k, тогда как остальными можно пренебречь. Фактически полная формальная теория почти не меняется при введении квантовой механики, но с ее помощью мы можем рассчитать величины f_k для данной электронной системы.

§ 2.4. Описание распространения электромагнитных волн с помощью интегральных уравнений*)

В начале настоящей главы был намечен подход, позволяющий описывать распространение электромагнитных волн с помощью интегральных уравнений, а в последующих разделах были введены вспомогательные величины, необходимые для этой цели. Теперь мы сформулируем эти интегральные уравнения и рассмотрим некоторые их следствия, что можно сделать, оставаясь полностью

*) Излагаемая здесь теория была развита главным образом в работах [16—19]. Полное изложение теории приведено в [20]. См. также [21, 22].

Некоторые обобщения этой теории, учитывающие нелинейное поведение среды (проявляющееся при исключительно сильных полях), рассмотрены в работе [23].

в рамках макроскопической теории, излагаемой на протяжении данной книги. В рамках этой теории интегральные уравнения эквивалентны уравнениям Максвелла и представляют собой математическое описание электромагнитных явлений с помощью взаимодействий на конечных расстояниях (им, конечно, необходимо время для распространения). Определенные преимущества такого подхода, который в ряде случаев оказывается мощнее обычного подхода, основанного на дифференциальных уравнениях, заключаются в том, что он связывает макроскопические явления с молекулярными, рассмотренными в предыдущем разделе.

Как мы уже говорили выше, можно считать, что молекулы, составляющие вещество, ведут себя в поле падающих волн подобно диполям. При этом все излучаемые диполями волны действуют на любой другой диполь с эффективной силой и определяют среднее измеряемое поле. Предположим, что диполи равномерно распределены по среде, и среднее значение их электрического момента в единице объема **P** будем рассматривать как основную величину. На самом же деле распределение молекул в среде никогда не бывает совершенно равномерным (т. е. имеются флуктуации плотности) и, следовательно, электрический момент отдельных частиц флуктуирует около среднего значения. Возникающие явления настоящая теория может объяснить, проводя расчеты несколько дальше, т. е. рассчитывая не только средние величины, но и их среднеквадратичные отклонения. Подобные расчеты важны для некоторых проблем, например для объяснения голубого цвета неба, впервые данного Рэлеем *). Но такое распространение теории здесь провести невозможно **).

2.4.1. Основное интегральное уравнение. Рассмотрим распространение электромагнитной волны в однородной изотропной немагнитной среде. Электрическое и магнитное поля $\mathbf{E}'_j$ и $\mathbf{H}'_j$, которые действуют на j-й диполь внутри среды, можно разделить на поля падающей волны $\mathbf{E}^{(i)}$, $\mathbf{H}^{(i)}$ (распространяющиеся со скоростью света в вакууме c) и вклад, создаваемый всеми диполями, т. е.

$$\mathbf{E}'_j=\mathbf{E}^{(i)}+\sum_l{}' \mathbf{E}_{jl},\qquad \mathbf{H}'_j=\mathbf{H}^{(i)}+\sum_l{}' \mathbf{H}_{jl}; \tag{1}$$

здесь суммирование распространяется на все диполи, кроме j-го. В точке $\mathbf{r}_j$, где расположен j-й диполь, поле от l-го диполя можно определить из формул (2.2.49) — (2.2.51), а именно

$$\mathbf{E}_{jl}=\operatorname{rot}\operatorname{rot}\frac{\mathbf{p}_l(t-R_{jl}/c)}{R_{jl}},\qquad \mathbf{H}_{jl}=\frac{1}{c}\operatorname{rot}\frac{\dot{\mathbf{p}}_l(t-R_{jl}/c)}{R_{jl}}, \tag{2}$$

где $\mathbf{p}_l(t)$ — момент l-го диполя, $R_{jl}=|\mathbf{r}_j-\mathbf{r}_l|$ и операция rot производится относительно координат x_j, y_j, z_j j-го диполя.

Как мы уже говорили, распределение можно с хорошим приближением считать непрерывным, т. е. момент диполей можно рассматривать как непрерывную функцию координат (и времени) $\mathbf{p}=\mathbf{p}(\mathbf{r},t)$. Концентрацию N также будем считать непрерывной функцией координат $N(\mathbf{r})$. В этом случае полный электрический дипольный момент **P** единицы объема определяется формулой (2.3.14), т. е.

$$\mathbf{P}=N\mathbf{p}=N\alpha\mathbf{E}'. \tag{3}$$

*) Позже мы будем вновь ссылаться на теорию Рэлея в § 13.5, где рассматривается рассеяние света сферическими частицами, находящимися в вакууме. Если частицы малы, зависимость интенсивности рассеянного излучения от длины волны совпадает с этой зависимостью в рассматриваемом случае спонтанных флуктуаций плотности в однородной среде (а именно, обратная пропорциональность четвертой степени длины волны). Однако, кроме этого результата, флуктуационная теория дает также зависимость интенсивности рассеянного излучения от флуктуаций плотности

**) Эта теория рассматривается, например, в § 81 книги [5].

По причинам, изложенным в § 2.3, мы пренебрегли вкладом в (3) силы, создаваемой магнитным полем. Поскольку допускается также, что вещество немагнитно (т. е $M=0$), в устанавливаемые ниже условия динамического равновесия не будет входить эффективное поле $\mathbf{H}'$.

Если подставить (2) в (1) и перейти к непрерывному распределению, используя (3), то получим *)

$$\mathbf{E}'(\mathbf{r},\ t)=\mathbf{E}^{(i)}+\int \operatorname{rot}\operatorname{rot} N\alpha\,\frac{\mathbf{E}'(\mathbf{r}',\ t-R/c)}{R}\,dV', \tag{4}$$

$$\mathbf{H}'(\mathbf{r},\ t)=\mathbf{H}^{(i)}+\frac{1}{c}\int \operatorname{rot} N\alpha\,\frac{\dot{\mathbf{E}}'(\mathbf{r}',\ t-R/c)}{R}\,dV', \tag{5}$$

где $R=|\mathbf{r}-\mathbf{r}'|$.

Если точка наблюдения $\mathbf{r}$ находится вне рассматриваемой среды, интеграл берется по всей среде. Если она расположена внутри среды, то необходимо вначале исключить небольшую область, занятую атомом; будем считать эту область небольшой сферой σ радиуса a. В конечном счете мы обычным образом перейдем к пределу $a\to 0$.

Уравнение (4) представляет собой интегро-дифференциальное уравнение относительно E'. Если решить его, то можно получить $\mathbf{H}'$ из (5). Эти два уравнения, по существу, эквивалентны уравнениям Максвелла для изотропных немагнитных веществ. Обобщение на магнитные среды можно провести с помощью второго вектора Герца.

2.4.2. Теорема погашения Эвальда — Озеена и строгий вывод формулы Лоренц — Лоренца. Уравнение (4) связывает довольно сложным образом эффективное электрическое поле с электрическим полем падающей волны. Это уравнение решается в явном виде лишь в специальных случаях. Тем не менее из него можно получить ряд основных результатов, таких, как формула Лоренц — Лоренца, законы преломления и отражения и формулы Френеля. Перед тем как показать это, выведем одно важное общее следствие решения.

Пусть Σ обозначает границу рассматриваемой среды. Для точек наблюдения, находящихся внутри среды, основное уравнение (4) можно записать в виде

$$\mathbf{P}=N\alpha(\mathbf{E}^{(i)}+\mathbf{E}^{(d)}), \tag{6}$$

где $\mathbf{E}^{(d)}$ — вклад от диполей, равный

$$\mathbf{E}^{(d)}=\int_{\sigma}^{\Sigma} \operatorname{rot}\operatorname{rot}\frac{\mathbf{P}(\mathbf{r}',\ t-R/c)}{R}\,dV'. \tag{7}$$

Здесь мы явно указали границу объема, по которому проводится интегрирование.

Будем считать, что поле $\mathbf{E}^{(i)}$ создается падающей монохроматической волной с угловой частотой ω, т. е.

$$\mathbf{E}^{(i)}=\mathbf{A}^{(i)}(\mathbf{r})\,e^{-i\omega t}. \tag{8}$$

В качестве пробного решения для $\mathbf{P}$ выберем волну, которая также монохроматична и имеет ту же частоту, но обладает другой скоростью распространения (скажем, c/n)

$$\mathbf{P}=(n^2-1)\,k_0^2\mathbf{Q}(\mathbf{r})\,e^{-i\omega t}, \tag{9}$$

где, как и раньше, $k_0=\omega/c$ и

$$\nabla^2\mathbf{Q}+n^2k_0^2\mathbf{Q}=0. \tag{10}$$

*) Мы должны суммировать поля, а не потенциалы, т. е. в (4) оператор rot rot должен стоять под знаком интеграла, а не вне его. В самом деле, поле, действующее на молекулу и выраженное в виде (3), получается суммированием всех индивидуальных полей. Наличие особенности у полей вблизи источников ведет к различию между суммой полей и полем, связанным с полным потенциалом (суммой потенциалов, связанных с индивидуальными полями). См. ниже соотношение (15).

Постоянный множитель $(n^2-1)k_0^2$ в (9) введен для упрощения последующих формул. Постоянную n следует рассматривать как неизвестную величину, причем ее определение является одной из главных задач настоящего анализа.

Предположим также, что внутри среды

$$\operatorname{div} \mathbf{Q} = 0. \tag{11}$$

На первый взгляд возможность решения основного интегро-дифференциального уравнения (6) в форме нашего пробного решения может показаться до некоторой степени странной, так как $\mathbf{E}^{(i)}$ представляет волну, распространяющуюся со скоростью света в вакууме c, тогда как $\mathbf{P}$, по предположению, распространяется со скоростью c/n. Однако будет показано, что поле диполей $\mathbf{E}^{(d)}$ можно выразить в виде суммы двух членов, один из которых удовлетворяет волновому уравнению в вакууме и в точности гасит падающую волну, тогда как другой удовлетворяет волновому уравнению для распространения со скоростью c/n. Поэтому можно считать, что падающая волна гасится в любой точке внутри среды в результате интерференции создаваемого ею поля с полем диполей; при этом появляется новая волна с иной скоростью распространения (а в общем случае и с иным направлением распространения). Найденный результат известен как *теорема погашения*; он был установлен вначале для кристаллических сред Эвальдом [24] и для изотропных сред Озееном [17].

Чтобы доказать теорему погашения, перепишем (6) в форме, не зависящей от времени

$$\mathbf{Q} = N\alpha \left\{ \frac{1}{(n^2-1)k_0^2} \mathbf{A}^{(i)} + \mathbf{A}^{(d)} \right\}, \tag{12}$$

где

$$\mathbf{E}^{(d)} = (n^2-1)k_0^2 \mathbf{A}^{(d)} e^{-i\omega t}, \qquad \mathbf{A}^{(d)} = \int_\sigma^\Sigma \operatorname{rot}\operatorname{rot} \mathbf{Q}(\mathbf{r}')\, G(R)\, dV' \tag{13}$$

и

$$G = \frac{\exp(ik_0 R)}{R}. \tag{14}$$

В приложении 5 показано, что если радиус сферы σ достаточно мал, то

$$\int_\sigma^\Sigma \operatorname{rot}\operatorname{rot} \mathbf{Q}(\mathbf{r}')\, G(R)\, dV' = \operatorname{rot}\operatorname{rot} \int_\sigma^\Sigma \mathbf{Q}(\mathbf{r}')\, G(R)\, dV' - \frac{8\pi}{3}\mathbf{Q}(\mathbf{r}). \tag{15}$$

Здесь функция G представляет сферическую волну в вакууме и поэтому удовлетворяет волновому уравнению

$$\nabla^2 G + k_0^2 G = 0. \tag{16}$$

Таким образом, из уравнений (10) и (16) следует, что член, стоящий под знаком интеграла в правой части (15), можно записать в виде

$$\mathbf{Q}G = \frac{1}{(n^2-1)k_0^2}(\mathbf{Q}\nabla^2 G - G\nabla^2 \mathbf{Q}). \tag{17}$$

При интегрировании с помощью теоремы Грина это дает

$$\int_\sigma^\Sigma \mathbf{Q}(\mathbf{r}')\, G(R)\, dV' = \\ = \frac{1}{(n^2-1)k_0^2}\left[\int_\Sigma \left\{ \mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'} \right\} dS' - \int_\sigma \left\{ \mathbf{Q}\frac{dG}{dR} - G\frac{\partial \mathbf{Q}}{\partial R} \right\} dS' \right], \tag{18}$$

где символ $\partial/\partial\nu'$ означает дифференцирование вдоль внешней нормали к границе Σ. Прямой расчет показывает (см. п. 8.3.1), что предел поверхностного

интеграла по σ при $a \to 0$ равен $-4\pi\mathbf{Q}(\mathbf{r})$ и, следовательно,

$$\int\limits_{\sigma}^{\Sigma} \mathbf{Q}(\mathbf{r}') G(R)\, dV' \to \frac{1}{(n^2-1)k_0^2} \left\{ \iint\limits_{\Sigma} \left(\mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'} \right) dS' + 4\pi\mathbf{Q} \right\}. \tag{19}$$

В пределе из (13), (15) и (19) получим

$$\mathbf{A}^{(d)} = \int\limits_{\sigma}^{\Sigma} \operatorname{rot}\operatorname{rot} \mathbf{Q}(\mathbf{r}') G(R)\, dV' =$$

$$= \frac{1}{(n^2-1)k_0^2} \left[4\pi \operatorname{rot}\operatorname{rot} \mathbf{Q} + \operatorname{rot}\operatorname{rot} \int\limits_{\Sigma} \left\{ \mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'} \right\} dS' \right] - \frac{8\pi}{3}\mathbf{Q}(\mathbf{r}). \tag{20}$$

Используем далее тождество

$$\operatorname{rot}\operatorname{rot}\mathbf{Q} = \operatorname{grad}\operatorname{div}\mathbf{Q} - \nabla^2\mathbf{Q}.$$

Первый член в правой части исчезает из-за условия (11), а второй, согласно (10), равен $n^2k_0^2\mathbf{Q}$. Следовательно, при подстановке их в (20) окончательно получим следующее выражение для поля диполей:

$$\mathbf{A}^{(d)} = \frac{4\pi}{3}\left(\frac{n^2+2}{n^2-1}\right)\mathbf{Q} + \frac{1}{(n^2-1)k_0^2} \operatorname{rot}\operatorname{rot} \int\limits_{\Sigma} \left\{ \mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'} \right\} dS'. \tag{21}$$

Согласно (10) первый член в правой части представляет волну, распространяющуюся со скоростью c/n, тогда как второй — волну, которая, подобно G, распространяется со скоростью света в вакууме c. Следовательно, как мы и ожидали, основное уравнение (12) разделяется на две группы членов, каждая из которых представляет волну, распространяющуюся со своей скоростью. Это возможно, только если каждая группа в нем равна нулю, и, следовательно,

$$\frac{4\pi}{3} N\alpha = \frac{n^2-1}{n^2+2} \tag{22}$$

и

$$\mathbf{A}^{(i)} + \operatorname{rot}\operatorname{rot} \int\limits_{\Sigma} \left\{ \mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'} \right\} dS' = 0. \tag{23}$$

Соотношение (23) выражает погашение падающей волны

$$\mathbf{E}^{(i)} = \mathbf{A}^{(i)} e^{-i\omega t}$$

в любой точке внутри среды, возникающее вследствие интерференции с частью поля диполей. Падающая волна заменяется другой, а именно волной вида

$$\mathbf{E}' = \frac{1}{N\alpha}\mathbf{P} = \frac{1}{N\alpha}(n^2-1)k_0^2\mathbf{Q}e^{-i\omega t} = \frac{4\pi}{3}(n^2+2)k_0^2\mathbf{Q}e^{-i\omega t}, \tag{24}$$

которая распространяется внутри среды со скоростью c/n. Величина n выражается в (22) через концентрацию N и поляризуемость α. В этом выражении мы узнаем *формулу Лоренц — Лоренца*, введенную предварительно в п. 2.3.3. Нужно отметить, что погашение падающей волны осуществляется исключительно диполями, расположенными на границе среды, а поляризация $\mathbf{P}$ возникает только вследствие взаимодействия соседних диполей.

В принципе общее решение интегро-дифференциального уравнения (23) можно получить, приведя его к системе двух связанных интегральных уравнений Фредгольма для $\mathbf{Q}$ и $\partial\mathbf{Q}/\partial\nu'$ на поверхности. Тогда величину $\mathbf{Q}$ в любой точке внутри среды можно получить, решая уравнение (10) при условии, что $\mathbf{Q}$ принимает указанные значения на границе. Это нетрудно сделать обычными методами, используя функции Грина или интегральную формулу Гельмгольца — Кирхгофа (см. уравнение (8.3.7)).

Мы должны теперь связать эффективное поле $\mathbf{E}'$ с обычным полем $\mathbf{E}$ теории Максвелла. Для этого запишем два определения электрического смещения $\mathbf{D}$, а именно

$$\mathbf{D}=\varepsilon\mathbf{E}=n^2\mathbf{E} \tag{25а}$$

и

$$\mathbf{D}=\mathbf{E}+4\pi\mathbf{P}. \tag{25б}$$

Исключая из них $\mathbf{D}$, найдем

$$\mathbf{E}=\frac{4\pi}{n^2-1}\,\mathbf{P}. \tag{26}$$

Теперь из (3) и (22) получим

$$\mathbf{E}'=\frac{1}{N\alpha}\mathbf{P}=\frac{4\pi}{3}\,\frac{n^2+2}{n^2-1}\mathbf{P}$$

и, следовательно, используя (26), имеем

$$\mathbf{E}'=\mathbf{E}+\frac{4\pi}{3}\mathbf{P}, \tag{27}$$

что согласуется с (2.3.12). Наконец, из (26) и (9) следует, что поле $\mathbf{E}$ *внутри среды* ($n\neq 1$) задается формулой

$$\mathbf{E}=4\pi k_0^2\mathbf{Q}e^{-i\omega t}. \tag{28}$$

Для нахождения электрического поля *вне среды* мы должны вернуться к уравнению (4). Интегралы, входящие в уравнения (4) и (5), теперь берутся по всей среде, так что оператор rot rot можно вынести из-под знака соответствующего интеграла. Далее, так как вне среды $\mathbf{D}=\mathbf{E}$, то $\mathbf{P}=0$ и, следовательно, согласно (27), мы формально получим $\mathbf{E}'=\mathbf{E}$. Вместо (6) теперь мы находим

$$\left.\begin{aligned}&\mathbf{E}=\mathbf{E}^{(i)}+\mathbf{E}^{(r)},\\ \text{где}\quad &\mathbf{E}^{(r)}=\operatorname{rot}\operatorname{rot}\int\frac{\mathbf{P}(\mathbf{r}',t-R/c)}{R}\,dV'.\end{aligned}\right\} \tag{29}$$

Для рассматриваемого здесь случая гармонической зависимости от времени получим, подставляя сюда выражение (9) и используя (17) и теорему Грина, соотношение

$$\left.\begin{aligned}&\mathbf{E}^{(r)}=\mathbf{A}^{(r)}e^{-i\omega t},\\ &\mathbf{A}^{(r)}=\operatorname{rot}\operatorname{rot}\int\limits_{\Sigma}\left\{\mathbf{Q}\frac{\partial G}{\partial\nu'}-G\frac{\partial\mathbf{Q}}{\partial\nu'}\right\}dS',\end{aligned}\right\} \tag{30}$$

а $\mathbf{Q}$ — та же функция на поверхности Σ, что и раньше.

До сих пор мы занимались только электрическим полем. Для нахождения магнитного поля нужно лишь подставить $\mathbf{E}'$ в (5) и взять интеграл. Это можно сделать методом, аналогичным приведенному выше, но такая операция сейчас несколько проще, так как оператор rot в (5) можно вынести из-под знака интеграла, что следует из результатов, изложенных в приложении 5. При сравнении получающихся выражений для $\mathbf{H}$ и для $\mathbf{E}$ мы увидим, что найденные решения согласуются с уравнениями Максвелла. Мы не будем приводить здесь эти расчеты, поскольку они довольно просты и не вносят каких-либо важных новых особенностей.

Чрезвычайно интересно отметить, что подход, основанный на физическом соображении о том, что поле в среде более естественно характеризовать поляризацией, чем вектором смещения, очень изящно, через интегродифференциальное уравнение (4), приводит к строгому выводу формулы Лорентц — Лоренца и теоремы погашения. Этот мощный метод до сих пор слабо использовался при

рассмотрении более частных проблем *); пример его применения будет дан в гл. 13.

2.4.3. Рассмотрение преломления и отражения плоской волны с помощью теоремы погашения Эвальда — Озеена. Применим теперь теорему погашения Эвальда — Озеена, выраженную формулой (23), к случаю плоской монохроматической волны, входящей в однородную среду, которая заполняет полупространство $z<0$. Покажем, что из этой теоремы вытекают законы преломления и отражения, а также формулы Френеля.

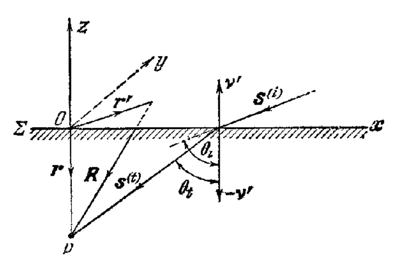


Рис. 2.4. Проникновение волны в однородную среду, рассматриваемую как система диполей.

Падающую волну запишем в форме

$$\mathbf{E}^{(i)} = \mathbf{A}^{(i)}(\mathbf{r})\, e^{-i\omega t} = \mathbf{A}_0^{(i)} \exp\{i\,[k_0(\mathbf{r}\cdot\mathbf{s}^{(i)}) - \omega t]\}, \quad (31)$$

где $k_0 = \omega/c$, $\mathbf{A}_0^{(i)}$ — постоянный вектор и $\mathbf{s}^{(i)}$ — единичный вектор в направлении распространения.

Выберем ось z так, чтобы она проходила через точку наблюдения p, которую мы вначале считаем находящейся внутри среды на расстоянии r от границы Σ (рис. 2.4). Ось x выбираем так, чтобы вектор $\mathbf{s}^{(i)}$ лежал в плоскости xz. Следовательно, обозначая угол падения через θ_i, мы получим

$$s_x^{(i)} = -\sin\theta_i, \qquad s_y^{(i)} = 0, \qquad s_z^{(i)} = -\cos\theta_i. \quad (32)$$

В соответствии с результатами предыдущего раздела, предположим, что прошедшая волна имеет ту же частоту, что и падающая, но другую скорость c/n, где n выражается через поляризуемость и плотность с помощью формулы Лоренц — Лоренца. В качестве пробного решения для прошедшей волны выберем плоскую волну, распространяющуюся в направлении единичного вектора $\mathbf{s}^{(t)}$, который, по предположению, лежит в плоскости xz:

$$s_x^{(t)} = -\sin\theta_t, \qquad s_y^{(t)} = 0, \qquad s_z^{(t)} = -\cos\theta_t. \quad (33)$$

Тогда выражение (28) примет вид

$$\mathbf{E} = 4\pi k_0^2 \mathbf{Q} e^{-i\omega t} = 4\pi k_0^2 \mathbf{Q}_0 \exp\{i\,[k_0 n(\mathbf{r}\cdot\mathbf{s}^{(t)}) - \omega t]\}, \quad (34)$$

где $\mathbf{Q}_0$— постоянный вектор, который, согласно (11), ортогонален $\mathbf{s}^{(t)}$.

Решение интегро-дифференциального уравнения легко получить, если рассматривать лишь точки наблюдения p, отстоящие от границы $z=0$ на расстояниях, больших по сравнению с длиной волны, т. е. если

$$2\pi r/\lambda = k_0 r \gg 1. \quad (35)$$

Это условие наложено лишь для упрощения расчетов и не соответствует ограничению физической обоснованности интегро-дифференциального уравнения.

Производные, входящие в интеграл по Σ, теперь имеют вид

$$\frac{\partial \mathbf{Q}(\mathbf{r}')}{\partial v'} = \frac{\partial}{\partial v'} \mathbf{Q}_0 \exp\{ink_0(\mathbf{r}'\cdot\mathbf{s}^{(t)})\} = ink_0\left(\frac{\partial \mathbf{r}'}{\partial v'}\cdot\mathbf{s}^{(t)}\right)\mathbf{Q}_0 \exp\{ink_0(\mathbf{r}'\cdot\mathbf{s}^{(t)})\}, \quad (36а)$$

$$\frac{\partial G(R)}{\partial v'} = \frac{\partial}{\partial v'}\frac{e^{ik_0R}}{R} = ik_0\frac{\partial R}{\partial v'}\, G\left(1 + \frac{i}{k_0R}\right). \quad (36б)$$

Последним членом в (36б) можно пренебречь из-за условия (35), и интеграл,

*) Этот метод был применен к исследованию распространения электромагнитных волн через слоистую среду в работе [25].

входящий в (23), записывается следующим образом:

$$\mathbf{J}=\int\limits_{\Sigma}\left(\mathbf{Q}\frac{\partial G}{\partial \mathbf{v}'}-G\frac{\partial \mathbf{Q}}{\partial \mathbf{v}'}\right)dS'=ik_0\mathbf{Q}_0\int\limits_{\Sigma}\left\{\frac{\partial R}{\partial \mathbf{v}'}-n\left(\frac{\partial \mathbf{r}'}{\partial \mathbf{v}'}\cdot\mathbf{s}^{(t)}\right)\right\}\frac{\exp\{ik_0[R+n(\mathbf{r}'\ \mathbf{s}^{(t)})]\}}{R}dS'. \tag{37}$$

Компоненты векторов **r**, **r′** и **R** равны (см. рис. 2.4)

$$\begin{array}{llll} \mathbf{r}: & 0, & 0, & -r, \\ \mathbf{r}': & x', & y', & 0, \\ \mathbf{R}: & -x', & -y', & -r, \end{array}$$

следовательно,

$$\left.\begin{aligned} R&=\sqrt{x'^2+y'^2+r^2}, & \mathbf{r}'\cdot\mathbf{s}^{(t)}&=-x'\sin\theta_t,\\ \frac{\partial R}{\partial \mathbf{v}'}&=\frac{r}{R}, & \frac{\partial \mathbf{r}'}{\partial \mathbf{v}'}\cdot\mathbf{s}^{(t)}&=s_z^{(t)}=-\cos\theta_t. \end{aligned}\right\} \tag{38}$$

Интеграл (37) после подстановки примет вид

$$\mathbf{J}=ik_0\mathbf{Q}_0\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\frac{1}{R}\left(\frac{r}{R}+n\cos\theta_t\right)\exp\{ik_0(R-nx'\sin\theta_t)\}\,dx'\,dy'. \tag{39}$$

Удобно ввести угол φ и соответствующий единичный вектор **s**, определяемые выражениями

$$n\sin\theta_t=\sin\varphi, \tag{40}$$

$$s_x=-\sin\varphi,\qquad s_y=0,\qquad s_z=-\cos\varphi. \tag{41}$$

Так как, по предположению, k_0r велико по сравнению с единицей, показатель степени экспоненциального члена в (39) также велик по сравнению с единицей и этот член будет быстро осциллировать и много раз менять знак, когда точка (x', y') будет пробегать область интегрирования. При таких условиях хорошее приближение к значению **J** получается путем применения следующей формулы, которая является следствием принципа стационарной фазы (см. приложение 3):

$$\iint\limits_{D} g(x', y')\exp[ik_0f(x', y')]\,dx'\,dy'\approx$$

$$\approx\frac{2\pi i}{k_0}\sum_j\frac{\sigma_j}{\sqrt{|\alpha_j\beta_j-\gamma_j^2|}}g(x'_j, y'_j)\exp[ik_0f(x'_j, y'_j)]. \tag{42}$$

Здесь (x'_j, y'_j) — точки в области интегрирования D, для которых f постоянна,

$$\alpha_j=\left(\frac{\partial^2 f}{\partial x'^2}\right)_{x'_j, y'_j},\qquad \beta_j=\left(\frac{\partial^2 f}{\partial y'^2}\right)_{x'_j, y'_j},\qquad \gamma_j=\left(\frac{\partial^2 f}{\partial x'\partial y'}\right)_{x'_j, y'_j} \tag{43а}$$

и

$$\sigma=\left\{\begin{array}{ll} +1 & \text{когда } \alpha_j\beta_j>\gamma_j^2,\quad \alpha_j>0,\\ -1, & \text{когда } \alpha_j\beta_j>\gamma_j^2,\quad \alpha_j<0,\\ -i, & \text{когда } \alpha_j\beta_j<\gamma_j^2 \end{array}\right\} \tag{43б}$$

Если использовать соотношения (40) и (41), то в данном случае мы получим

$$g=\frac{1}{R}\left(\frac{r}{R}+\frac{\sin\varphi}{\sin\theta_t}\cos\theta_t\right),\qquad f=R-x'\sin\varphi, \tag{44}$$

и, следовательно,

$$\frac{\partial f}{\partial x'}=\frac{x'}{\sqrt{x'^2+y'^2+r^2}}-\sin\varphi,\qquad \frac{\partial f}{\partial y'}=\frac{y'}{\sqrt{x'^2+y'^2+r^2}}.$$

Для постоянства f необходимо, чтобы

$$\frac{\partial f}{\partial x'}=\frac{\partial f}{\partial y'}=0.$$

Следовательно, как легко показать (отметим, что при вычислении частных производных необходимо брать положительный квадратный корень), f будет постоянна, только если

$$x'=x_1', \qquad y'=y_1',$$

где

$$x_1'=r\,\mathrm{tg}\,\varphi, \qquad y_1'=0. \tag{45}$$

В этой точке выражения (43) и (44) принимают следующие значения:

$$\alpha_1=\left(\frac{\partial^2 f}{\partial x'^2}\right)_{x_1',\,y_1'}=\frac{\cos^3\varphi}{r}, \quad \beta_1=\left(\frac{\partial^2 f}{\partial y'^2}\right)_{x_1',\,y_1'}=\frac{\cos\varphi}{r}, \quad \gamma_1=0, \tag{46а}$$

$$\sigma_1=1, \tag{46б}$$

$$R_1=(R)_{x_1',\,y_1'}=\frac{r}{\cos\varphi}, \tag{46в}$$

$$f(x_1',\ y_1')=r\cos\varphi=\mathbf{r}\cdot\mathbf{s}, \tag{46г}$$

$$g(x_1',\ y_1')=\frac{\cos\varphi}{r}\,\frac{\sin(\varphi+\theta_t)}{\sin\theta_t}. \tag{46д}$$

Теперь (42) примет вид

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\frac{1}{R}\left(\frac{r}{R}+n\cos\theta_t\right)\exp\{ik_0(R-nx'\sin\theta_t)\}\,dx'\,dy'=$$
$$=\frac{2\pi i}{k_0}\,\frac{\sin(\varphi+\theta_t)}{\cos\varphi\sin\theta_t}\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})],$$

так что (39) равно

$$\mathbf{J}=\int\limits_{\Sigma}\left\{\mathbf{Q}\frac{\partial G}{\partial\nu'}-G\frac{\partial\mathbf{Q}}{\partial\nu'}\right\}dS'=-2\pi\frac{\sin(\varphi+\theta_t)}{\cos\varphi\sin\theta_t}\,\mathbf{Q}_0\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})]. \tag{47}$$

Подставив теперь это выражение в интегро-дифференциальное уравнение (23), которое выражает погашение падающей волны, получим

$$\mathrm{rot}\,\mathbf{Q}_0\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})]=ik_0(\mathbf{s}\times\mathbf{Q}_0)\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})],$$
$$\mathrm{rot}\,\mathrm{rot}\,\mathbf{Q}_0\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})]=-k_0^2\mathbf{s}\times(\mathbf{s}\times\mathbf{Q}_0)\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})].$$

Следовательно, (23) перейдет в

$$\mathbf{A}_0^{(i)}\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s}^{(i)})]=-2\pi k_0^2\frac{\sin(\varphi+\theta_t)}{\cos\varphi\sin\theta_t}\,\mathbf{s}\times(\mathbf{s}\times\mathbf{Q}_0)\exp\,[ik_0(\mathbf{r}\cdot\mathbf{s})]. \tag{48}$$

Это соотношение должно тождественно выполняться для всех точек $\mathbf{r}$ на границе. Следовательно,

$$\mathbf{s}=\mathbf{s}^{(i)} \tag{49}$$

и

$$\mathbf{A}_0^{(i)}=-2\pi k_0^2\frac{\sin(\varphi+\theta_t)}{\cos\varphi\sin\theta_t}\{\mathbf{s}^{(i)}\times(\mathbf{s}^{(i)}\times\mathbf{Q}_0)\}. \tag{50}$$

Уравнение (49) выражает *закон преломления* (1.5.8). В самом деле, он означает, что

$$\varphi=\theta_i, \tag{51}$$

или, в соответствии с (40),

$$n\sin\theta_t=\sin\theta_i. \tag{52}$$

Более того, соотношения (41) и (32) показывают, что вектор $\mathbf{s}^{(t)}$ лежит в плоскости, определяемой вектором $\mathbf{s}^{(i)}$ и нормалью к границе Σ.

Уравнение (50) связывает амплитуды падающей волны (31) и прошедшей волны (34). Обозначая амплитуду (векторную) прошедшей волны через $\mathbf{T}_0$,

$$\mathbf{T}_0 = 4\pi k_0^2 \mathbf{Q}_0, \tag{53}$$

и раскрывая тройное векторное произведение, получим вместо (50)

$$\mathbf{A}_0^{(i)} = \frac{1}{2}\frac{\sin(\theta_i+\theta_t)}{\cos\theta_i \sin\theta_t}\left[\mathbf{T}_0 - \mathbf{s}^{(i)}(\mathbf{s}^{(i)}\cdot\mathbf{T}_0)\right]. \tag{54}$$

Пусть $A_\perp$, $A_\parallel$ и $T_\perp$, $T_\parallel$ — составляющие $\mathbf{A}_0^{(i)}$ и $\mathbf{T}_0$ в направлениях, перпендикулярном и параллельном плоскости падения. Тогда, вспоминая, что вектор $\mathbf{A}_0^{(i)}$ ортогонален $\mathbf{s}^{(i)}$, а $\mathbf{T}_0$ ортогонален $\mathbf{s}^{(t)}$ и что $\mathbf{s}^{(i)}$ и $\mathbf{s}^{(t)}$ составляют друг с другом угол $(\theta_i - \theta_t)$, найдем из (54)

$$A_\perp = \frac{1}{2}\frac{\sin(\theta_i+\theta_t)}{\cos\theta_i \sin\theta_t} T_\perp, \tag{55а}$$

$$A_\parallel = \frac{1}{2}\frac{\sin(\theta_i+\theta_t)\cos(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t} T_\parallel. \tag{55б}$$

В этих соотношениях мы узнаём *формулы Френеля для преломления* (см. (1.5.20а)).

Наконец, рассмотрим случай, когда точка наблюдения находится вне среды ($z>0$). Расчеты совершенно аналогичны, только теперь в соответствующих формулах $z' = -r$ следует заменить на $z' = r$. Это эквивалентно замене s_z на $-s_z$ в (46г), т. е. замене $\varphi = \theta_i$ на $\varphi = \theta_r$, где

$$\theta_r = \pi - \theta_i. \tag{56}$$

Вместо единичного вектора $\mathbf{s} = \mathbf{s}^{(i)}$ введем теперь единичный вектор $\mathbf{s}^{(r)}$ с компонентами

$$s_x^{(r)} = -\sin\theta_r = -\sin\theta_i, \quad s_y^{(r)} = 0, \quad s_z^{(r)} = -\cos\theta_r = \cos\theta_i. \tag{57}$$

и получим вместо (47) соотношение

$$\mathbf{J}' = \int\limits_\Sigma \left\{ \mathbf{Q}\frac{\partial G}{\partial \nu'} - G\frac{\partial \mathbf{Q}}{\partial \nu'}\right\} dS' = -2\pi \frac{\sin(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t}\mathbf{Q}_0 \cdot \exp\left[ik_0(\mathbf{r}\cdot\mathbf{s}^{(r)})\right]. \tag{58}$$

Из выражения (30) для не зависящей от времени части отраженной волны находим

$$\mathbf{A}_0^{(r)} \exp\left[ik_0(\mathbf{r}\cdot\mathbf{s}^{(r)})\right] = 2\pi k_0^2 \frac{\sin(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t}\left\{\mathbf{s}^{(r)}\times(\mathbf{s}^{(r)}\times\mathbf{Q}_0)\right\}\exp\left[ik_0(\mathbf{r}\cdot\mathbf{s}^{(r)})\right]. \tag{59}$$

Выражение (59) представляет плоскую волну, распространяющуюся в направлении, определяемом единичным вектором $\mathbf{s}^{(r)}$, причем это направление связано с направлением $\mathbf{s}^{(i)}$ падающей волны соотношением (56). Соотношение (56) выражает *закон отражения* в согласии с (1.5.7). Амплитуда $\mathbf{A}_0^{(r)}$ отраженной волны, выраженная через амплитуду прошедшей, записывается в виде

$$\mathbf{A}_0^{(r)} = -\frac{1}{2}\frac{\sin(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t}\left[\mathbf{T}_0 - \mathbf{s}^{(r)}(\mathbf{s}^{(r)}\cdot\mathbf{T}_0)\right]. \tag{60}$$

Обозначая через $R_\perp$ и $R_\parallel$ составляющие $\mathbf{A}_0^{(r)}$ в направлениях, перпендикулярном и параллельном плоскости падения, и используя (55), получим

$$R_\perp = -\frac{1}{2}\frac{\sin(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t} T_\perp = -\frac{\sin(\theta_i-\theta_t)}{\sin(\theta_i+\theta_t)} A_\perp \tag{61а}$$

и

$$R_\parallel = -\frac{1}{2}\frac{\sin(\theta_i-\theta_t)\cos(\theta_i-\theta_t)}{\cos\theta_i \sin\theta_t} T_\parallel = \frac{\operatorname{tg}(\theta_i-\theta_t)}{\operatorname{tg}(\theta_i+\theta_t)} A_\parallel. \tag{61б}$$

Соотношения (61) совпадают с *формулами Френеля для отражения* (1.5.21а),

ЛИТЕРАТУРА

1. S. A. Schelkunoff, Electromagnetic Waves, Van Nostrand, New York, 1943, p. 382.
2. H. S. Green, E. Wolf, Proc. Phys. Soc. **A66**, 1129 (1953).
3. A. Nisbet, Proc. Roy. Soc. **A231**, 251 (1955).
4. M. Born, Natural Philosophy of Cause and Chance, Clarendon Press, Oxford, 1949.
5. M. Born, Optik, Springer, Berlin, 1933. (М. Борн, Оптика, ГНТИУ, Харьков — Киев, 1937.)
6. H. Hertz, Ann. d. Physik, **36**, 1 (1889).
7. A. Righi, Nuovo Cimento **2**, 104 (1901).
8. F. Kottler, Ann. d. Physik **71**, 462 (1923); A. Nisbet, Proc. Roy. Soc. **A231**, 250 (1955)
9. W. H. McCrea, Proc. Roy. Soc. **A240**, 447 (1957).
10. H. A. Lorentz, Wiedem. Ann. **9**, 641 (1880).
11. L. Lorenz, Wiedem. Ann. **11**, 70 (1881).
12. R. Clausius, Mechanische Wärmetheorie, Braunschweig, 2nd ed. 1879, Vol. **2**, p. 62.
13. O. F. Mossotti, Mem. Soc. Sci. Modena **14**, 49 (1850).
14. J. Koch, Nov. Act. Soc Upsala **2**, 61 (1909).
15. L. Cauchy, Bull. des. Sci. Math. **14**, 9 (1830); Sur la dispersion de la lumière, Nouv. exerc. de math. 1836.
16. W. Esmarch, Ann. d. Physik, **42**, 1257 (1913).
17. C. W. Oseen, Ann. d. Physik **48**, 1 (1915).
18. W. Boethe, Beiträge zur Theorie der Brechung und Reflexion, Dissertation, Berlin, 1914.
19. W. Boethe, Ann. d. Physik **64**, 693 (1921).
20. R. Lundblad, Univ. Arsskrift, Upsala, 1920.
21. C. G. Darwin, Trans. Cambr. Phil. Soc. **23**, 137 (1924).
22. H. Hoek, Algemeene theorie der optische activiteit van isotrope media, Thesis, Leiden, 1939.
23. N. Bloembergen, P. S. Pershan, Phys. Rev. **128**, 619 (1962).
24. P. P. Ewald, Dissertation, München, 1912; Ann. d. Physik **49**, 1 (1916).
25. C. W. Oseen, Ann. d. Physik **48**, 1 (1915).
26. D. R. Hartree, Proc. Cambr. Phil. Soc. **25**, 97 (1929).

ГЛАВА 3

ОСНОВЫ ГЕОМЕТРИЧЕСКОЙ ОПТИКИ

§ 3.1. Приближение очень коротких длин волн

Электромагнитное поле видимого света характеризуется очень быстрыми колебаниями (с частотой порядка 10^{14} *сек*$^{-1}$), или, что то же, малостью длин волн (порядка 10^{-5} *см*). Поэтому можно ожидать, что для законов распространения видимого света получится хорошее первое приближение, если полностью пренебречь конечностью длин волн. Оказывается, что такая процедура справедлива при решении многих оптических задач; более того, физические явления, которые не описываются этой приближенной теорией (так называемые дифракционные явления, рассматривающиеся в гл. 8), можно обнаружить лишь с помощью очень тонких экспериментов.

Раздел оптики, в котором пренебрегают конечностью длин волн, что соответствует предельному переходу $\lambda_0 \to 0$, называется *геометрической оптикой* *), поскольку в этом приближении оптические законы можно сформулировать на языке геометрии. В рамках геометрической оптики считается, что энергия распространяется вдоль определенных кривых (световых лучей). Физическую модель пучка световых лучей можно получить, если пропустить свет от источника пренебрежимо малого размера через небольшое отверстие в непрозрачном экране. Свет, выходящий из отверстия, заполняет область, граница которой (край пучка лучей) кажется на первый взгляд резкой. Однако более тщательное исследование показывает, что интенсивность света около границы изменяется хотя и быстро, но непрерывно, от нуля в области тени до максимального значения в освещенной области. Это изменение не является монотонным, а носит периодический характер, что приводит к появлению светлых и темных полос, называемых дифракционными. Размер области, в которой происходит это быстрое изменение интенсивности света, порядка длины волны. Следовательно, если длина волны пренебрежимо мала по сравнению с размерами отверстия, то можно говорить о пучке световых лучей с резкой границей **). При уменьшении размеров отверстия до величины, сравнимой с длиной волны, возникают эффекты, для объяснения которых требуется более тонкое исследование. Однако если ограничиться рассмотрением предельного случая пренебрежимо малых длин волн, то на размер отверстия не накладывается никакого ограничения и можно считать, что из отверстия исчезающе малых размеров выходит бесконечно тонкий пучок света — световой луч. Будем показано, что изменение поперечного сечения пучка световых лучей служит мерой изменения интенсивности света. Кроме того, мы увидим, что с каждым лучом можно связать состояние поляризации и исследовать его изменение вдоль луча.

Далее будет показано, что для коротких длин волн общий характер поля такой же, как и в случае плоской волны; более того, законы преломления и отражения, установленные для плоской волны, падающей на плоскую границу, остаются в приближении геометрической оптики справедливыми и при более общих условиях. Следовательно, если на резкую границу (например, поверх-

*) История развития геометрической оптики описана в работах [1—3].

**) Тот факт, что граница становится резкой в пределе $\lambda_0 \to 0$, был впервые установлен *Кирхгофом* [4]. См. также [5, 6].

ность линзы) падает луч света, то он разделяется на отраженный и преломленный лучи, а изменение состояния поляризации и отражательную, и пропускательную способности можно вычислить из соответствующих формул для плоских волн.

Из приведенных выше рассуждений следует, что для достаточно коротких длин волн полное объяснение оптических явлений можно получить из геометрических соображений путем определения пути световых лучей и соответствующих им интенсивности и состояния поляризации. Сформулируем теперь соответствующие законы, переходя в уравнениях Максвелла к пределу *) $\lambda_0 \to 0$.

3.1.1. Вывод уравнения эйконала. Рассмотрим гармонически меняющееся со временем поле общего вида

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r})\, e^{-i\omega t}, \qquad \mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0(\mathbf{r})\, e^{-i\omega t} \tag{1}$$

в непроводящей изотропной среде. $\mathbf{E}_0$ и $\mathbf{H}_0$ являются комплексными векторными функциями положения и, как показано в п. 1.4.3, вещественные части стоящих справа выражений (1) описывают физические поля.

Комплексные векторы $\mathbf{E}_0$ и $\mathbf{H}_0$ удовлетворяют уравнениям Максвелла в форме, не зависящей от времени, которые получаются при подстановке выражений (1) в уравнения (1.1.1)—(1.1.4). В областях, свободных от токов и зарядов ($\mathbf{j} = \rho = 0$), уравнения Максвелла имеют вид

$$\operatorname{rot} \mathbf{H}_0 + ik_0\varepsilon \mathbf{E}_0 = 0, \tag{2}$$

$$\operatorname{rot} \mathbf{E}_0 - ik_0\mu \mathbf{H}_0 = 0, \tag{3}$$

$$\operatorname{div} \varepsilon \mathbf{E}_0 = 0, \tag{4}$$

$$\operatorname{div} \mu \mathbf{H}_0 = 0. \tag{5}$$

При выводе этих уравнений были использованы материальные соотношения $\mathbf{D} = \varepsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, и, как и раньше, соотношения $k_0 = \omega/c = 2\pi/\lambda_0$, где λ_0 — длина волны в вакууме.

Ранее было показано, что однородная плоская волна, распространяющаяся в среде с показателем преломления $n = \sqrt{\varepsilon\mu}$ в направлении, определяемом единичным вектором $\mathbf{s}$, описывается следующими выражениями:

$$\mathbf{E}_0 = \mathbf{e} \exp(ik_0 n(\mathbf{s}\cdot\mathbf{r})), \qquad \mathbf{H}_0 = \mathbf{h} \exp(ik_0 n(\mathbf{s}\cdot\mathbf{r})), \tag{6}$$

где $\mathbf{e}$ и $\mathbf{h}$ — постоянные векторы, в общем случае комплексные. Для монохроматического поля электрического диполя в вакууме было найдено (см. § 2.2)

$$\mathbf{E}_0 = \mathbf{e} \exp(ik_0 r), \qquad \mathbf{H}_0 = \mathbf{h} \exp(ik_0 r), \tag{7}$$

где r — расстояние от диполя. Здесь векторы $\mathbf{e}$ и $\mathbf{h}$ уже не являются постоянными, но на достаточно больших расстояниях от диполя ($r \gg \lambda_0$) и при соответствующей нормировке дипольного момента они не зависят от k_0.

На основании этого разумно предположить, что в областях, расположенных на расстояниях многих длин волн от источника, поля более общего типа можно представить в виде

$$\mathbf{E}_0 = \mathbf{e}(\mathbf{r}) \exp(ik_0 \mathcal{S}(\mathbf{r})), \qquad \mathbf{H}_0 = \mathbf{h}(\mathbf{r}) \exp(ik_0 \mathcal{S}(\mathbf{r})), \tag{8}$$

где $\mathcal{S}(\mathbf{r})$ — «оптический путь» — вещественная скалярная функция положения, а $\mathbf{e}(\mathbf{r})$ и $\mathbf{h}(\mathbf{r})$ — векторные функции положения, в общем случае комплексные**).

---

*) Зоммерфельд и Рунге [7], используя идею Дебая, впервые показали, что основное уравнение геометрической оптики (уравнение эйконала (15б)) можно вывести из скалярного волнового уравнения при $\lambda_0 \to 0$. Обобщение этого вывода, учитывающее векторный характер электромагнитного поля, проведено в работах [8—14].

**) $\mathbf{e}$ и $\mathbf{h}$ обязательно должны быть комплексными, если необходимо учесть все возможные состояния поляризации. Вещественные $\mathbf{e}$ и $\mathbf{h}$, согласно (1.4.75), соответствуют линейно поляризованному свету.

Подстановка (8) в качестве пробного решения в уравнения Максвелла дает несколько соотношений между $\mathbf{e}$, $\mathbf{h}$ и $\mathcal{S}$. Будет показано, что в случае больших k_0 (малых λ_0) из этих соотношений вытекает требование, чтобы величина $\mathcal{S}$ удовлетворяла некоторому дифференциальному уравнению, не зависящему от векторных амплитуд $\mathbf{e}$ и $\mathbf{h}$.

Используя хорошо известные векторные тождества, получим из (8)

$$\operatorname{rot} \mathbf{H}_0 = (\operatorname{rot} \mathbf{h} + ik_0 \operatorname{grad} \mathcal{S} \times \mathbf{h}) \exp(ik_0\mathcal{S}), \tag{9}$$

$$\operatorname{div} \mu \mathbf{H}_0 = (\mu \operatorname{div} \mathbf{h} + \mathbf{h}\cdot\operatorname{grad} \mu + ik_0\mu\mathbf{h}\cdot\operatorname{grad} \mathcal{S}) \exp(ik_0\mathcal{S}) \tag{10}$$

и аналогичные выражения для $\operatorname{rot} \mathbf{E}_0$ и $\operatorname{div} \varepsilon \mathbf{E}_0$. Следовательно, уравнения (2) — (5) примут вид

$$\operatorname{grad} \mathcal{S} \times \mathbf{h} + \varepsilon\mathbf{e} = -\frac{1}{ik_0} \operatorname{rot} \mathbf{h}, \tag{11}$$

$$\operatorname{grad} \mathcal{S} \times \mathbf{e} - \mu\mathbf{h} = -\frac{1}{ik_0} \operatorname{rot} \mathbf{e}, \tag{12}$$

$$\mathbf{e}\cdot\operatorname{grad} \mathcal{S} = -\frac{1}{ik_0} (\mathbf{e}\cdot\operatorname{grad} \ln \varepsilon + \operatorname{div} \mathbf{e}), \tag{13}$$

$$\mathbf{h}\cdot\operatorname{grad} \mathcal{S} = -\frac{1}{ik_0} (\mathbf{h}\cdot\operatorname{grad} \ln \mu + \operatorname{div} \mathbf{h}). \tag{14}$$

Нас интересует решение, соответствующее случаю очень больших k_0. Поэтому правыми частями уравнений (11)—(14) можно пренебречь, если выражения, которые умножаются на $1/ik_0$, не будут чрезвычайно большими. В этом случае уравнения (11)—(14) запишутся следующим образом:

$$\operatorname{grad} \mathcal{S} \times \mathbf{h} + \varepsilon\mathbf{e} = 0, \tag{11a}$$

$$\operatorname{grad} \mathcal{S} \times \mathbf{e} - \mu\mathbf{h} = 0, \tag{12a}$$

$$\mathbf{e}\cdot\operatorname{grad} \mathcal{S} = 0, \tag{13a}$$

$$\mathbf{h}\cdot\operatorname{grad} \mathcal{S} = 0. \tag{14a}$$

Ограничимся изучением лишь уравнений (11а) и (12а), так как (13а) и (14а) получаются из них, если первые скалярно умножить на $\operatorname{grad} \mathcal{S}$. Уравнения (11а) и (12а) можно рассматривать как совместную систему шести однородных линейных скалярных уравнений относительно декартовых компонент e_x, h_x, ... векторов $\mathbf{e}$ и $\mathbf{h}$. Эта система имеет нетривиальное решение лишь в случае выполнения условия совместности (равенство нулю соответствующего определителя). Последнее легко получить, исключая $\mathbf{e}$ или $\mathbf{h}$ из (11а) и (12а). Подставляя $\mathbf{h}$ из (12а) в (11а), находим

$$\frac{1}{\mu} \{(\mathbf{e}\cdot\operatorname{grad} \mathcal{S}) \operatorname{grad} \mathcal{S} - \mathbf{e} (\operatorname{grad} \mathcal{S})^2\} + \varepsilon\mathbf{e} = 0.$$

Первый член обращается в нуль на основании (13а), и поскольку $\mathbf{e}$ не равно нулю во всем пространстве, находим

$$(\operatorname{grad} \mathcal{S})^2 = n^2, \tag{15a}$$

или в явном виде

$$\left(\frac{\partial\mathcal{S}}{\partial x}\right)^2 + \left(\frac{\partial\mathcal{S}}{\partial y}\right)^2 + \left(\frac{\partial\mathcal{S}}{\partial z}\right)^2 = n^2(x, y, z), \tag{15б}$$

где, как и ранее, $n = \sqrt{\varepsilon\mu}$ — показатель преломления. Функцию $\mathcal{S}$ часто называют *эйконалом* *), а соотношение (15б) — *уравнением эйконала*; оно является

---

*) Термин «эйконал» (от греческого слова εἰκών — изображение) был введен в 1895 г. Брунсом для обозначения некоторых связанных функций (см. стр. 138), однако в дальнейшем он стал применяться в более широком смысле.

основным уравнением геометрической оптики *). Поверхности $\mathcal{S}(\mathbf{r}) = \text{const}$ называются *геометрическими волновыми поверхностями или геометрическими волновыми фронтами* **).

Уравнение эйконала было выведено нами из уравнений Максвелла (дифференциальных уравнений первого порядка), однако его можно получить и из волновых уравнений (уравнений второго порядка) для векторов электрического или магнитного полей. Для этого следует подставить выражения (1) и (8) в волновое уравнение (1.2.5), и тогда после простых преобразований получим

$$\mathsf{K}(\mathbf{e}, \mathcal{S}, n) + \frac{1}{ik_0}\mathsf{L}(\mathbf{e}, \mathcal{S}, n, \mu) + \frac{1}{(ik_0)^2}\mathsf{M}(\mathbf{e}, \varepsilon, \mu) = 0, \tag{16}$$

где

$$\mathsf{K}(\mathbf{e}, \mathcal{S}, n) = \{n^2 - (\operatorname{grad}\mathcal{S})^2\}\,\mathbf{e},$$

$$\mathsf{L}(\mathbf{e}, \mathcal{S}, n, \mu) = \{\operatorname{grad}\mathcal{S}\cdot\operatorname{grad}\ln\mu - \nabla^2\mathcal{S}\}\,\mathbf{e} - \\ - 2\{\mathbf{e}\cdot\operatorname{grad}\ln n\}\operatorname{grad}\mathcal{S} - 2(\operatorname{grad}\mathcal{S}\cdot\operatorname{grad})\,\mathbf{e},$$

$$\mathsf{M}(\mathbf{e}, \varepsilon, \mu) = \operatorname{rot}\mathbf{e}\times\operatorname{grad}\ln\mu - \nabla^2\mathbf{e} - \operatorname{grad}(\mathbf{e}\cdot\operatorname{grad}\ln\varepsilon).$$

Соответствующее уравнение, содержащее $\mathbf{h}$ и получающееся после подстановки выражения для $\mathbf{H}$ в (1.2.6) (либо, проще, используя то обстоятельство, что уравнения Максвелла остаются неизменными при одновременной замене $\mathbf{E}$ на $\mathbf{H}$ и ε на $-\mu$), имеет вид

$$\mathsf{K}(\mathbf{h}, \mathcal{S}, n) + \frac{1}{ik_0}\mathsf{L}(\mathbf{h}, \mathcal{S}, n, \varepsilon) + \frac{1}{(ik_0)^2}\mathsf{M}(\mathbf{h}, \mu, \varepsilon) = 0. \tag{17}$$

В случае достаточно больших k_0 вторыми и третьими членами в (16) и (17) обычно можно пренебречь; тогда $\mathsf{K} = 0$, откуда сразу же следует уравнение эйконала. Позднее будет показано, что члены в (16) и (17), содержащие $1/ik_0$ в первой степени, также имеют физическое истолкование.

Можно показать, что во многих важных случаях векторы $\mathbf{E}_0$ и $\mathbf{H}_0$ можно разложить в асимптотические ряды вида ***)

$$\mathbf{E}_0 = \exp(ik_0\mathcal{S})\sum_{m\geqslant 0}\frac{\mathbf{e}^{(m)}}{(ik_0)^m}, \qquad \mathbf{H}_0 = \exp(ik_0\mathcal{S})\sum_{m\geqslant 0}\frac{\mathbf{h}^{(m)}}{(ik_0)^m}, \tag{18}$$

где $\mathbf{e}^{(m)}$ и $\mathbf{h}^{(m)}$ — функции координат, а $\mathcal{S}$ — та же функция, что и раньше ****). Геометрической оптике соответствуют первые члены этих разложений.

3.1.2. Световые лучи и закон интенсивности в геометрической оптике. Из соотношений (8), (1.4.54) и (1.4.55) следует, что усредненные по времени плотности электрической $\langle w_e\rangle$ и магнитной $\langle w_m\rangle$ энергии записываются

---

*) Уравнение эйконала можно считать также характеристическим уравнением волновых уравнений (5) и (6) из § 1.2 для $\mathbf{E}$ и $\mathbf{H}$. Оно дает строгое описание распространения разрывов решений этих уравнений. Однако в геометрической оптике интересуются не распространением разрывов, а решениями, гармонически (или почти гармонически) меняющимися со временем. В приложении 6 показана формальная эквивалентность этих двух интерпретаций.

Уравнение эйконала можно рассматривать также как уравнение Гамильтона — Якоби для вариационной задачи $\delta\int n\,ds = 0$, впервые поставленной применительно к оптике Ферма (см. п. 3.3.2 и приложение 1).

**) В дальнейшем в тех местах, где это не вызовет недоразумений, мы будем опускать прилагательное «геометрический».

***) Здесь мы предполагаем, что через каждую точку проходит только один геометрический волновой фронт. В некоторых случаях, например, при наличии в среде отражающих препятствий, через каждую точку может проходить несколько волновых фронтов. Тогда результирующее поле получается сложением нескольких рядов подобного типа.

****) Теория таких асимптотических разложений была впервые развита главным образом в работе Лунеберга [16]. См. также [17а], [18] и [22]. Исчерпывающий обзор теории приведен в [17].

в виде

$$\langle w_e \rangle = \frac{\varepsilon}{16\pi}\,\mathbf{e}\cdot\mathbf{e}^*, \qquad \langle w_m \rangle = \frac{\mu}{16\pi}\,\mathbf{h}\cdot\mathbf{h}^*. \tag{19}$$

Подставляя сюда $\mathbf{e}^*$ из (11а) и $\mathbf{h}$ из (12а), получим

$$\langle w_e \rangle = \langle w_m \rangle = \frac{1}{16\pi}\,[\mathbf{e},\ \mathbf{h}^*,\ \operatorname{grad}\mathcal{S}], \tag{20}$$

где квадратные скобки обозначают смешанное произведение. Следовательно, *в приближении геометрической оптики усредненные по времени плотности электрической и магнитной энергии равны друг другу.*

Среднее по времени значение вектора Пойнтинга можно получить с помощью (8) и (1.4.52); имеем

$$\langle \mathbf{S} \rangle = \frac{c}{8\pi}\operatorname{Re}(\mathbf{e}\times\mathbf{h}^*).$$

Используя (12а), получим

$$\langle \mathbf{S} \rangle = \frac{c}{8\pi\mu}\,\{(\mathbf{e}\cdot\mathbf{e}^*)\operatorname{grad}\mathcal{S} - (\mathbf{e}\cdot\operatorname{grad}\mathcal{S})\,\mathbf{e}^*\}.$$

На основании (13а) последний член в этом выражении равен нулю; выражая затем $\mathbf{e}\cdot\mathbf{e}^*$ через $\langle w_e \rangle$ и используя соотношение Максвелла $\varepsilon\mu = n^2$, находим

$$\langle \mathbf{S} \rangle = \frac{2c}{n^2}\langle w_e \rangle \operatorname{grad}\mathcal{S}. \tag{21}$$

Поскольку $\langle w_e \rangle = \langle w_m \rangle$ член $2\langle w_e \rangle$ определяет среднюю по времени плотность полной энергии $\langle w \rangle$ ($\langle w \rangle = \langle w_e \rangle + \langle w_m \rangle$). Выражение же $(\operatorname{grad}\mathcal{S})/n$ на основании уравнения эйконала является единичным вектором (скажем, $\mathbf{s}$), т. е.

$$\mathbf{s} = \frac{\operatorname{grad}\mathcal{S}}{n} = \frac{\operatorname{grad}\mathcal{S}}{|\operatorname{grad}\mathcal{S}|}; \tag{22}$$

тогда из (21) следует, что вектор $\mathbf{s}$ направлен вдоль усредненного вектора Пойнтинга. Полагая, как и раньше, $c/n = v$, выражение (21) можно переписать в виде

$$\langle \mathbf{S} \rangle = v\langle w \rangle \mathbf{s}. \tag{23}$$

Следовательно, *направление усредненного по времени вектора Пойнтинга совпадает с нормалью к геометрическому волновому фронту, а абсолютная его величина равна произведению средней плотности энергии на скорость $v = c/n$.* Этот результат аналогичен соотношению (1.4.9) для случая плоских волн и свидетельствует о том, что *в приближении геометрической оптики средняя плотность энергии распространяется со скоростью $v = c/n$.*

Геометрические световые лучи можно теперь определить как траектории, ортогональные к геометрическим волновым фронтам $\mathcal{S} = \text{const}$. Мы будем приписывать этим линиям направление, полагая, что оно совпадает в каждой точке с направлением усредненного вектора Пойнтинга *). Если радиус-вектор $\mathbf{r}(s)$ точки P, расположенной на луче, рассматривать как функцию длин s дуги луча, то $d\mathbf{r}/ds = \mathbf{s}$, и уравнение луча запишется в виде

$$n\frac{d\mathbf{r}}{ds} = \operatorname{grad}\mathcal{S}. \tag{24}$$

Из (13а) и (14а) видно, что *векторы электрического и магнитного полей в каждой точке ортогональны лучу.*

---

*) Такое определение световых лучей справедливо лишь для изотропных сред Позже, в гл. 14 мы увидим, что в анизотропной среде нормаль к волновому фронту в общем случае не совпадает с направлением вектора Пойнтинга.

Смысл уравнения (24) можно пояснить следующим образом. Рассмотрим два соседних волновых фронта $\mathcal{S} = \text{const}$ и $\mathcal{S} + d\mathcal{S} = \text{const}$ (рис. 3.1). Тогда

$$\frac{d\mathcal{S}}{ds} = \frac{d\mathbf{r}}{ds} \cdot \operatorname{grad} \mathcal{S} = n. \tag{25}$$

Следовательно, расстояние ds между точками пересечения нормали с этими волновыми фронтами обратно пропорционально показателю преломления, т. е. прямо пропорционально v.

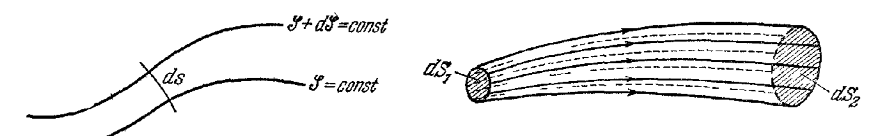


Рис. 3.1. Чертеж, поясняющий смысл соотношения $n\mathbf{s} = \operatorname{grad} \mathcal{S}$

Рис. 3.2. К выводу закона интенсивности в геометрической оптике

Интеграл $\int\limits_C n\,ds$ вдоль кривой C называется *оптической длиной* этой кривой. Показывая квадратными скобками оптическую длину *луча*, соединяющего точки P_1 и P_2, получим

$$[P_1P_2] = \int\limits_{P_1}^{P_2} n\,ds = \mathcal{S}(P_2) - \mathcal{S}(P_1). \tag{26}$$

Как мы установили, средняя плотность энергии распространяется вдоль луча со скоростью $v = c/n$, поэтому

$$n\,ds = \frac{c}{v}\,ds = c\,dt,$$

где dt — время прохождения энергией расстояния ds вдоль луча; следовательно,

$$[P_1P_2] = c \int\limits_{P_1}^{P_2} dt, \tag{27}$$

т. е. *оптическая длина* $[P_1P_2]$ *равна произведению скорости света в вакууме на время распространения света от* P_1 *к* P_2.

Интенсивность света I была определена нами как абсолютное значение от среднего по времени вектора Пойнтинга. Поэтому из (23) следует, что

$$I = |\langle \mathbf{S} \rangle| = v \langle w \rangle, \tag{28}$$

а закон сохранения (1.4.57) дает

$$\operatorname{div}(I\mathbf{s}) = 0. \tag{29}$$

Для того чтобы понять смысл этого соотношения, рассмотрим узкую трубку, образованную световыми лучами, выходящими из элемента dS_1 волнового фронта $\mathcal{S}(\mathbf{r}) = a_1$ (где a_1 — постоянная), и обозначим через dS_2 соответствующий элемент, который пересекают эти лучи на другом волновом фронте $\mathcal{S}(\mathbf{r}) = a_2$ (рис. 3.2). Интегрируя (29) по объему трубки и применяя теорему Гаусса, получим

$$\int I\mathbf{s} \cdot \boldsymbol{\nu}\, dS = 0, \tag{30}$$

где $\boldsymbol{\nu}$ — внешняя нормаль к поверхности трубки. Поскольку

$$\mathbf{s} \cdot \boldsymbol{\nu} = \begin{cases} 1 & \text{на } dS_2, \\ -1 & \text{на } dS_1, \\ 0 & \text{на остальной поверхности,} \end{cases}$$

(30) сводится к выражению

$$I_1 dS_1 = I_2 dS_2, \qquad (31)$$

где I_1 и I_2 — интенсивности на dS_1 и dS_2 соответственно. Следовательно, *величина $I\,dS$ остается постоянной вдоль трубки лучей.* Это соотношение выражает *закон интенсивности в геометрической оптике*

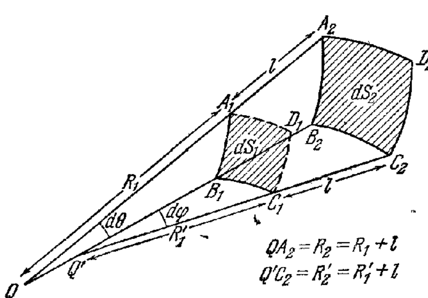


Рис. 3.3. К выводу закона интенсивности в геометрической оптике для прямолинейных лучей

Позже будет показано, что в однородной среде световые лучи имеют вид прямых линий. Закон интенсивности в этом случае можно представить в несколько иной форме. Предположим сначала, что dS_1, а следовательно, и dS_2 ограничены отрезками линий кривизны (рис. 3.3). Если R_1 и R_1' — главные радиусы кривизны отрезков A_1B_1 и B_1C_1, то

$$A_1B_1 = R_1 d\theta, \qquad B_1C_1 = R_1' d\varphi,$$

где $d\theta$ и $d\varphi$ — углы, под которыми видны A_1B_1 и B_1C_1 из соответствующих центров кривизны Q и Q'. Следовательно,

$$dS_1 = A_1B_1 \cdot B_1C_1 = R_1 R_1' \, d\theta \, d\varphi. \qquad (32)$$

Аналогичное выражение получается для элемента dS_2, который выделяется пучком лучей, прошедших через dS_1, из другого волнового фронта семейства, т. е.

$$dS_2 = A_2B_2 \cdot B_2C_2 = R_2 R_2' \, d\theta \, d\varphi. \qquad (33)$$

Если l — расстояние между dS_1 и dS_2, измеренное вдоль луча, то

$$R_2 = R_1 + l, \qquad R_2' = R_1' + l,$$

и окончательно получим

$$\frac{I_2}{I_1} = \frac{dS_1}{dS_2} = \frac{R_1 R_1'}{R_1 R_2'} = \frac{R_1 R_1'}{(R_1 + l)(R_1' + l)}. \qquad (34)$$

Если площадки dS_1 и dS_2 ограничены произвольными кривыми, то формула (34) все равно остается справедливой. В этом легко убедиться, если разбить площадки на большое число элементов, ограниченных линиями кривизны, а затем просуммировать вклады от всех элементов.

Если $R_1 \ll l$, $R_1' \ll l$, то (34) сводится к

$$\frac{I_2}{I_1} = \frac{R_1 R_1'}{l^2}. \qquad (35)$$

Этой формулой иногда пользуются при изучении рассеяния света.

Величина $1/RR'$, обратная произведению двух главных радиусов кривизны, называется *гауссовой* (или второй) *кривизной* поверхности. Из (34) следует, что *в любой точке прямолинейного луча интенсивность пропорциональна гауссовой кривизне волнового фронта, проходящего через эту точку.* В частности, если все (прямолинейные) лучи имеют одну общую точку, то волновые фронты имеют вид сферических поверхностей с центром в этой точке, тогда $R_1 = R_1'$, $R_2 = R_2'$, и мы получим (опуская индексы) *закон обратного квадрата расстояния*, т. е.

$$I = \frac{\text{const}}{R^2}. \qquad (36)$$

Возвращаясь к общему случаю произвольного пучка лучей (искривленных или прямых), мы можем, воспользовавшись функцией $\mathcal{S}$, записать в явном виде соотношения для изменения интенсивности вдоль каждого луча. Подставляя $\mathbf{s}$ из (22) в (29) и используя тождества $\operatorname{div} u\mathbf{v} = u \operatorname{div} \mathbf{v} + \mathbf{v}\cdot\operatorname{grad} u$, $\operatorname{div}\operatorname{grad} = \nabla^2$, получим

$$\frac{I}{n}\nabla^2\mathcal{S} + \operatorname{grad}\mathcal{S}\cdot\operatorname{grad}\frac{I}{n} = 0.$$

Это можно переписать следующим образом:

$$\nabla^2\mathcal{S} + \operatorname{grad}\mathcal{S}\cdot\operatorname{grad}\ln\frac{I}{n} = 0. \tag{37}$$

Введем теперь оператор

$$\frac{\partial}{\partial\tau} = \operatorname{grad}\mathcal{S}\cdot\operatorname{grad}, \tag{38}$$

где τ — параметр, характеризующий положение вдоль луча. Тогда (37) примет вид

$$\frac{\partial}{\partial\tau}\ln\frac{I}{n} = -\nabla^2\mathcal{S},$$

и после интегрирования находим

$$I = n\exp\left\{-\int^{\tau}\nabla^2\mathcal{S}\,d\tau\right\}.$$

Однако из (38), (15) и (25) следует, что

$$d\tau = \frac{d\mathcal{S}}{(\operatorname{grad}\mathcal{S})^2} = \frac{1}{n^2}d\mathcal{S} = \frac{1}{n}ds, \tag{39}$$

и мы получим окончательное выражение для отношения интенсивностей в двух произвольных точках луча в виде

$$\frac{I_2}{I_1} = \frac{n_2}{n_1}\exp\left\{-\int_{\mathcal{S}_1}^{\mathcal{S}_2}\frac{\nabla^2\mathcal{S}}{n^2}\,d\mathcal{S}\right\} = \frac{n_2}{n_1}\exp\left\{-\int_{s_1}^{s_2}\frac{\nabla^2\mathcal{S}}{n}\,ds\right\}; \tag{40}$$

здесь интегрирование проводится вдоль луча *).

3.1.3. Распространение векторных амплитуд. Мы видели, что в случае достаточно коротких длин волн распространение энергии можно описать с помощью простой гидродинамической модели, полностью характеризующейся вещественной скалярной функцией $\mathcal{S}$, которая служит решением уравнения эйконала (15) По традиции считается, что в геометрической оптике рассматривается именно такая приближенная картина распространения энергии, в которой использовались понятия луча и волновых фронтов. Другими словами, поляризация света не рассматривалась Это, без сомнения, объясняется тем, что простые законы геометрической оптики, относящиеся к лучам и волновым фронтам, были известны из экспериментов задолго до появления электромагнитной теории света Однако можно, и с нашей точки зрения вполне естественно, расширить рамки геометрической оптики, включив в нее некоторые геометрические законы, связанные с распространением векторных амплитуд $\mathbf{e}$ и $\mathbf{h}$. Эти законы легко получить из волновых уравнений (16) и (17).

Поскольку функция $\mathcal{S}$ удовлетворяет уравнению эйконала, $\mathbf{K} = 0$, и мы видим, что при достаточно больших k_0 (малых λ_0) в уравнениях (16) и (17) остаются лишь члены, содержащие $\mathbf{L}$ Следовательно, в рассматриваемом приближении векторные амплитуды и эйконал связаны между собой соотноше-

*) Клейн [20] показал, что отношение интенсивностей (40) можно выразить через интеграл, содержащий главные радиусы кривизны соответствующих волновых фронтов Формула Клейна является естественным обобщением соотношения (34) для неоднородной среды. См. также [17].

ниями $\mathsf{L} = 0$. Если снова воспользоваться оператором $\partial/\partial\tau$, определенным (38), то уравнения $\mathsf{L} = 0$ примут вид

$$\frac{\partial \mathbf{e}}{\partial\tau}+\frac{1}{2}\left(\nabla^2\mathcal{S}-\frac{\partial \ln\mu}{\partial\tau}\right)\mathbf{e}+(\mathbf{e}\cdot\operatorname{grad}\ln n)\operatorname{grad}\mathcal{S}=0, \tag{41}$$

$$\frac{\partial \mathbf{h}}{\partial\tau}+\frac{1}{2}\left(\nabla^2\mathcal{S}-\frac{\partial \ln\varepsilon}{\partial\tau}\right)\mathbf{h}+(\mathbf{h}\cdot\operatorname{grad}\ln n)\operatorname{grad}\mathcal{S}=0. \tag{42}$$

Это и есть искомые *уравнения переноса*, описывающие изменения $\mathbf{e}$ и $\mathbf{h}$ вдоль каждого луча Чтобы лучше понять их физический смысл, необходимо отдельно рассмотреть изменения этих векторов по величине и направлению.

Умножим (41) скалярно на $\mathbf{e}^*$ и к полученному уравнению прибавим комплексно сопряженное. Это дает

$$\frac{\partial}{\partial\tau}(\mathbf{e}\cdot\mathbf{e}^*)+\left(\nabla^2\mathcal{S}-\frac{\partial\ln\mu}{\partial\tau}\right)\mathbf{e}\cdot\mathbf{e}^*=0. \tag{43}$$

Учитывая тождество $\operatorname{div} u\mathbf{v} = u\operatorname{div}\mathbf{v}+\mathbf{v}\cdot\operatorname{grad} u$, разность $(\nabla^2\mathcal{S}-\partial(\ln\mu)/\partial\tau)$ можно записать следующим образом

$$\nabla^2\mathcal{S}-\frac{\partial\ln\mu}{\partial\tau}=\nabla^2\mathcal{S}-\operatorname{grad}\mathcal{S}\cdot\operatorname{grad}\ln\mu=\mu\operatorname{div}\left(\frac{1}{\mu}\operatorname{grad}\mathcal{S}\right). \tag{44}$$

Интегрируя (43) вдоль луча, получим для отношения величин $\mathbf{e}\cdot\mathbf{e}^*$ в любых двух точках луча соотношение

$$\frac{(\mathbf{e}\cdot\mathbf{e}^*)_2}{(\mathbf{e}\cdot\mathbf{e}^*)_1}=\exp\left(-\int\limits_{\mathcal{S}_1}^{\mathcal{S}_2}\mu\operatorname{div}\left(\frac{1}{\mu}\operatorname{grad}\mathcal{S}\right)d\tau\right)=$$
$$=\exp\left(-\int\limits_{s_1}^{s_2}\sqrt{\frac{\mu}{\varepsilon}}\operatorname{div}\left(\frac{1}{\mu}\operatorname{grad}\mathcal{S}\right)ds\right). \tag{45}$$

Это соотношение можно представить и в форме

$$\left(\frac{\mathbf{e}\cdot\mathbf{e}^*}{\mu}\right)_2=\left(\frac{\mathbf{e}\cdot\mathbf{e}^*}{\mu}\right)_1\exp\left\{-\int\limits_{s_1}^{s_2}\frac{\nabla^2\mathcal{S}}{n}\,ds\right\}, \tag{45a}$$

которая получается, если (43) записать в виде

$$\frac{\partial}{\partial\tau}\left[\ln\left(\frac{\mathbf{e}\cdot\mathbf{e}^*}{\mu}\right)\right]=-\nabla^2\mathcal{S}$$

и интегрировать вдоль луча Фактически соотношение (45а) является лишь иной формой выражения (40) для изменения интенсивности и следует из него, если использовать выражение

$$I=\frac{2c}{n}\langle w_e\rangle=\frac{c\varepsilon}{8\pi n}(\mathbf{e}\cdot\mathbf{e}^*)$$

и формулу Максвелла $\varepsilon\mu=n^2$

Аналогичным образом получим

$$\frac{(\mathbf{h}\cdot\mathbf{h}^*)_2}{(\mathbf{h}\cdot\mathbf{h}^*)_1}=\exp\left\{-\int\limits_{s_1}^{s_2}\sqrt{\frac{\varepsilon}{\mu}}\operatorname{div}\left(\frac{1}{\varepsilon}\operatorname{grad}\mathcal{S}\right)ds\right\}. \tag{46}$$

Рассмотрим теперь изменение комплексных единичных векторов

$$\mathbf{u}=\frac{\mathbf{e}}{\sqrt{\mathbf{e}\cdot\mathbf{e}^*}},\qquad \mathbf{v}=\frac{\mathbf{h}}{\sqrt{\mathbf{h}\cdot\mathbf{h}^*}} \tag{47}$$

вдоль каждого луча Подстановка в (41) дает

$$\frac{\partial\mathbf{u}}{\partial\tau}+\frac{1}{2}\left[\frac{\partial(\ln\mathbf{e}\cdot\mathbf{e}^*)}{\partial\tau}+\nabla^2\mathcal{S}-\frac{\partial\ln\mu}{\partial\tau}\right]\mathbf{u}+(\mathbf{u}\cdot\operatorname{grad}\ln n)\operatorname{grad}\mathcal{S}=0.$$

Согласно (43) выражение в квадратных скобках равно нулю, и мы имеем

$$\frac{d\mathbf{u}}{d\tau}=n\frac{d\mathbf{u}}{ds}=-(\mathbf{u}\cdot\operatorname{grad}\ln n)\operatorname{grad}\mathcal{S}, \tag{48}$$

и аналогично

$$\frac{d\mathbf{v}}{d\tau} = n\frac{d\mathbf{v}}{ds} = -(\mathbf{v}\,\mathrm{grad}\ln n)\,\mathrm{grad}\,\mathcal{S}. \qquad (49)$$

Эти соотношения и представляют законы изменения $\mathbf{u}$ и $\mathbf{v}$ вдоль луча *). В частности, *для однородной среды* ($n = \mathrm{const}$) (48) и (49) сводятся к выражениям $d\mathbf{u}/ds = d\mathbf{v}/ds = 0$, т. е. *векторы* $\mathbf{u}$ *и* $\mathbf{v}$ *в этом случае остаются постоянными вдоль каждого луча.*

Наконец, отметим, что для гармонической по времени однородной плоской волны в однородной среде $\mathcal{S} = n\mathbf{s}\cdot\mathbf{r}$, а $\mathbf{e}$, $\mathbf{h}$, ε и μ постоянны; следовательно, в уравнении (16) $\mathsf{K} = \mathsf{L} = \mathsf{M} = 0$. Поэтому такая волна (независимо от ее частоты) строго подчиняется законам геометрической оптики.

3.1.4. Обобщения геометрической оптики и пределы ее применимости. Выводы предыдущих разделов относились к строго монохроматическому полю. Такое поле, которое можно рассматривать как фурье-компоненту произвольного поля, создает гармонический осциллятор или набор подобных осцилляторов с одинаковой частотой.

В оптике обычно имеют дело с источником, излучающим свет в узком, но конечном диапазоне частот. Такой источник можно рассматривать как набор, большого числа гармонических осцилляторов, частоты которых попадают в указанный диапазон. Для вычисления интенсивности света в какой-то точке P необходимо просуммировать все поля, созданные каждым осциллятором (элементом источника), т. е.

$$\mathbf{E} = \sum_n \mathbf{E}_n, \qquad \mathbf{H} = \sum_n \mathbf{H}_n. \qquad (50)$$

Тогда интенсивность определяется (в вещественном представлении) соотношением

$$I(P) = |\langle\mathbf{S}\rangle| = \frac{c}{4\pi}|\langle\mathbf{E}\times\mathbf{H}\rangle| = \frac{c}{4\pi}\left|\sum_{n,\,m}\langle\mathbf{E}_n\times\mathbf{H}_m\rangle\right| =$$
$$= \frac{c}{4\pi}\left|\sum_n\langle\mathbf{E}_n\times\mathbf{H}_n\rangle + \sum_{n\neq m}\langle\mathbf{E}_n\times\mathbf{H}_m\rangle\right|. \qquad (51)$$

Во многих оптических задачах можно допустить, что вторая сумма в (51) равна нулю (в этих случаях говорят, что поля *некогерентны*); тогда

$$I(P) = \frac{c}{4\pi}\left|\sum_n\langle\mathbf{E}_n\times\mathbf{H}_n\rangle\right| = \left|\sum_n\langle\mathbf{S}_n\rangle\right|, \qquad (52)$$

где $\mathbf{S}_n$ — вектор Пойнтинга, соответствующий n-му элементу источника. Сейчас мы еще не в состоянии выяснить условия, при которых оправдано пренебрежение вторым членом в формуле (51), однако мы это сделаем позже при рассмотрении частичной когерентности (см. гл. 10).

Пусть δS — небольшая часть волнового фронта, соответствующего какому-то определенному элементу источника. Через δS проходят трубки лучей, исходящих от каждого элемента источника, центральные лучи этих трубок заполняют конус с телесным углом $\delta\Omega$ (рис. 3.4). Если угол раствора конуса достаточ-

---

*) Выражения (48) и (49) интересно интерпретируются в неевклидовой геометрии. Если рассмотреть соответствующее неевклидово пространство, в котором элемент длины дается соотношением

$$ds' = n\,ds = n\sqrt{dx^2 + dy^2 + dz^2},$$

то геометрические световые лучи в этом пространстве совпадают с геодезическими линиями, а соотношения (48) и (49), как можно показать, свидетельствуют о том, что векторы $\mathbf{u}$ и $\mathbf{v}$ перемещаются вдоль каждого луча параллельно (в смысле Леви-Чивита) самим себе (см. [15], [17], стр. 180–183 и [21]).

но мал, то можно пренебречь зависимостью S_n от направления и записать (52) в виде

$$I(P)=\sum_n |\langle \mathbf{S}_n\rangle| = \sum_n I_n. \qquad (53)$$

Теперь положим, что число элементов (осцилляторов) настолько велико, что их распределение без существенной ошибки можно считать непрерывным. Вклад от каждого элемента бесконечно мал, однако суммарный эффект конечен. В этом случае сумма (интеграл) пропорциональна $\delta\Omega$, т. е. $I(P)=B\delta\Omega$,

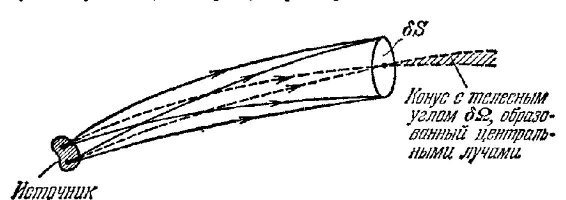


Рис. 3.4. К выводу закона интенсивности в геометрической оптике для некогерентного источника конечных размеров.

и полный (усредненный по времени) поток энергии δF, проходящий через элемент δS в единицу времени, равен

$$\delta F = B\delta\Omega\,\delta S. \qquad (54)$$

Последняя формула играет важную роль в фотометрии и будет использована позже.

Рассмотрим теперь кратко пределы применимости геометрической оптики. Уравнение эйконала было получено в предположении, что членами, стоящими в правых частях соотношений (11) и (12), можно пренебречь. Если допустить что безразмерные величины ε, μ и $|\operatorname{grad}\mathcal{S}|$ порядка единицы, то, как мы видим, пренебрежение указанными выше членами оправдано, когда изменения $\mathbf{e}$ и $\mathbf{h}$ на расстояниях, сравнимых с длиной волны, малы по сравнению с самими величинами $\mathbf{e}$ и $\mathbf{h}$. Это условие нарушается, например, на границах тени, так как там интенсивность (а следовательно, $\mathbf{e}$ и $\mathbf{h}$) резко меняется. Нельзя также ожидать, что геометрическая оптика даст правильное описание полей вблизи точек, где интенсивность имеет резкий максимум (например, в фокусе, см. § 8 8).

Уравнения переноса (41) и (42) для комплексных векторных амплитуд $\mathbf{e}$ и $\mathbf{h}$ были выведены в предположении, что функция $\mathcal{S}$ удовлетворяет уравнению эйконала, а члены $\lambda_0|\mathbf{M}(\mathbf{e}, \varepsilon, \mu|$ и $\lambda_0|\mathbf{M}(\mathbf{h}, \mu, \varepsilon)|$ малы по сравнению с $|\mathbf{L}(\mathbf{e}, \mathcal{S}, n, \mu)|$ и $|\mathbf{L}(\mathbf{h}, \mathcal{S}, n, \varepsilon)|$ соответственно. Эти предположения накладывают некоторые дополнительные ограничения не только на первые, но и на вторые производные от $\mathbf{e}$ и $\mathbf{h}$. Соответствующие условия довольно громоздки, и мы их рассматривать не будем.

Можно, конечно, получить более точное приближение, оставляя в разложениях для полей некоторые члены более высоких порядков *). Однако практическая ценность такой процедуры для решения задач инструментальной оптики весьма сомнительна, поскольку чем ближе мы подходим к особым областям, тем больше членов в разложениях надо оставлять, а в точках, представляющих наибольший интерес (в фокусе или на каустической поверхности),

*) Келлер [23] предположил, что вклад от членов высшего порядка можно изучить с помощью модели являющейся обобщением обычной геометрической оптики В этой теории вводится понятие *дифрагировавшего луча*, подчиняющегося обобщенному принципу Ферма Каждому такому лучу соответствует поле, и предполагается, что последнее подчиняется тем же общим законам распространения, что и поля в геометрической оптике. Были рассмотрены некоторые приложения этой теории [24, 25] (см также [17]).

эти разложения, как правило, расходятся. Для изучения распределения интенсивности в таких областях более эффективны методы, которые будут рассмотрены в главах, посвященных дифракции.

Наконец подчеркнем, что простота геометрической оптики связана в основном с тем, что обычно в *каждой точке* поле представляет собой плоскую волну. В оптическом диапазоне частот области, в которых простая геометрическая модель оказывается несправедливой, встречаются весьма редко; фактически в большинстве оптических задач эта модель дает по крайней мере хорошее нулевое приближение для более тонкого исследования.

§ 3.2. Общие свойства лучей

3.2.1. Дифференциальное уравнение для световых лучей. Световые лучи были определены как траектории, ортогональные к геометрическим волновым фронтам $\mathcal{S}(x, y, z) = \text{const}$, и мы показали, что если $\mathbf{r}$ есть радиус-вектор произвольной точки луча, а s — длина луча, отсчитываемая от этой точки на нем, то

$$n\frac{d\mathbf{r}}{ds} = \text{grad}\,\mathcal{S}. \tag{1}$$

Уравнение (1) описывает поведение лучей с помощью функции $\mathcal{S}$, однако из него легко получить дифференциальное уравнение, характеризующее лучи непосредственно показателем преломления $n(\mathbf{r})$.

Дифференцируя (1) по s, получим

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \frac{d}{ds}(\text{grad}\,\mathcal{S}) = \frac{d\mathbf{r}}{ds}\cdot\text{grad}\,(\text{grad}\,\mathcal{S}) =$$
$$= \frac{1}{n}\text{grad}\,\mathcal{S}\cdot\text{grad}\,(\text{grad}\,\mathcal{S}) = \frac{1}{2n}\text{grad}\,(\text{grad}\,\mathcal{S})^2 = \frac{1}{2n}\text{grad}\,n^2.$$

При этих преобразованиях использовано соотношение (3.1.15). Таким образом,

$$\frac{d}{ds}\left(n\frac{d\mathbf{r}}{ds}\right) = \text{grad}\,n. \tag{2}$$

Последнее соотношение представляет собой векторную форму дифференциальных уравнений для световых лучей. В частности, *в однородной среде* $n = \text{const}$ и (2) принимает вид

$$\frac{d^2\mathbf{r}}{ds^2} = 0,$$

откуда

$$\mathbf{r} = s\mathbf{a} + \mathbf{b}; \tag{3}$$

здесь $\mathbf{a}$ и $\mathbf{b}$ — постоянные векторы. Соотношение (3) — это векторное уравнение прямой линии, направленной по $\mathbf{a}$ и проходящей через точку $\mathbf{r} = \mathbf{b}$. Следовательно, *в однородной среде световые лучи являются прямыми линиями.*

В качестве примера, представляющего известный интерес, расмотрим поведение лучей в среде, обладающей центральной симметрией, т. е. в среде, показатель преломления которой зависит только от расстояния r до фиксированной точки O:

$$n = n(r). \tag{4}$$

Подобные условия приближенно выполняются в земной атмосфере, если учитывается кривизна Земли.

Рассмотрим изменение вектора $\mathbf{r}\times[n(r)\mathbf{s}]$ вдоль луча. Имеем

$$\frac{d}{ds}(\mathbf{r}\times n\mathbf{s}) = \frac{d\mathbf{r}}{ds}\times n\mathbf{s} + \mathbf{r}\times\frac{d}{ds}(n\mathbf{s}). \tag{5}$$

Поскольку $d\mathbf{r}/ds = \mathbf{s}$, первый член в правой части (5) обращается в нуль. Второй член можно на основании (2) переписать в виде $\mathbf{r} \times \mathbf{grad}\, n$. Далее из (4) получим

$$\mathrm{grad}\, n = \frac{\mathbf{r}}{r}\frac{dn}{dr};$$

таким образом, второй член в правой части (5) также обращается в нуль. Следовательно,

$$\mathbf{r} \times n\mathbf{s} = \mathrm{const}. \tag{6}$$

Отсюда следует, что все лучи являются плоскими кривыми, лежащими в плоскости, проходящей через начало координат, и вдоль каждого луча выполняется условие

$$nr \sin \varphi = \mathrm{const}, \tag{7}$$

где φ — угол между радиусом-вектором $\mathbf{r}$ и касательной в точке P (рис. 3.5). Так как величина $r \sin \varphi$ равна расстоянию от начала координат до касательной, выражение (7) можно записать в виде

$$nd = \mathrm{const}. \tag{8}$$

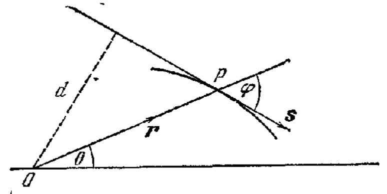


Рис. 3.5. К выводу формулы Бугера $nd = \mathrm{const}$ для лучей, распространяющихся в сферически симметричной среде.

Это соотношение иногда называют *формулой Бугера*; она является аналогом известной формулы динамики, выражающей закон сохранения углового момента частицы, движущейся под действием центральной силы.

Чтобы получить в явном виде уравнения световых лучей в сферически симметричной среде, вспомним из элементарной геометрии, что если (r, θ) — полярные координаты, то угол φ между радиусом-вектором точки P на плоской кривой и касательной в этой точке дается соотношением (см., например, [26])

$$\sin \varphi = \frac{r(\theta)}{\sqrt{r^2(\theta) + \left(\frac{dr}{d\theta}\right)^2}}. \tag{9}$$

Из (7) и (9) найдем

$$\frac{dr}{d\theta} = \frac{r}{c}\sqrt{n^2r^2 - c^2}, \tag{10}$$

где c — постоянная. Тогда уравнение лучей в сферически симметричной среде можно записать в виде

$$\theta = c \int^{r} \frac{dr}{r\sqrt{n^2r^2 - c^2}}. \tag{11}$$

Вернемся теперь к обсуждению общего случая и рассмотрим *вектор кривизны* луча, т. е. вектор

$$\mathbf{K} = \frac{d\mathbf{s}}{ds} = \frac{1}{\rho}\boldsymbol{\nu}, \tag{12}$$

длина которого $1/\rho$ равна величине, обратной радиусу кривизны; $\boldsymbol{\nu}$ — единичный вектор главной нормали в произвольной точке луча.

Из соотношений (2) и (12) следует, что

$$n\mathbf{K} = \mathrm{grad}\, n - \frac{dn}{ds}\mathbf{s}. \tag{13}$$

Это соотношение показывает, что *градиент показателя преломления лежит в соприкасающейся плоскости луча.*

Умножив (13) скалярно на $\mathbf{K}$ и воспользовавшись (12), получим

$$|\mathbf{K}| = \frac{1}{\rho} = \boldsymbol{\nu} \cdot \mathrm{grad}\ln n. \tag{14}$$

Поскольку величина ρ всегда положительна, отсюда следует, что вдоль главной нормали показатель преломления возрастает, т. е. *луч «загибается» в область с бóльшим показателем преломления* (рис. 3.6).

3.2.2. **Законы преломления и отражения.** До сих пор предполагалось, что показатель преломления n — непрерывная функция. Рассмотрим теперь поведение лучей, пересекающих поверхность, разделяющую две однородные среды с различными показателями преломления. Зоммерфельд и Рунге [7] показали, что его легко установить с помощью рассуждений, сходных с рассуждениями, которые проводились при выводе граничных условий для векторов полей на поверхности раздела (см. п. 1.1.3).

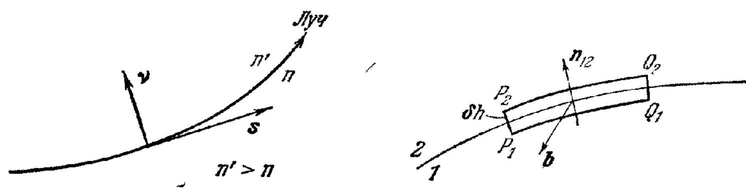


Рис. 3.6. Искривление луча в неоднородной среде.

Рис. 3.7. К выводу законов преломления и отражения.

Учитывая тождество rot grad $\equiv 0$, находим, что в соответствии с (1) вектор $n\mathbf{s} = n\,d\mathbf{r}/ds$, называемый иногда *лучевым вектором*, удовлетворяет соотношению

$$\operatorname{rot} n\mathbf{s} = 0. \tag{15}$$

Как и в п. 1.1.3, заменим поверхность раздела T переходным слоем, в котором величины ε, μ и n меняются быстро, но непрерывно от своих значений около T с одной стороны поверхности до значений около T с другой ее стороны. Далее рассмотрим плоский элемент поверхности, стороны которого P_1Q_1 и P_2Q_2 параллельны, а P_1P_2 и Q_1Q_2 перпендикулярны к T (рис. 3.7). Если обозначить через $\mathbf{b}$ единичный вектор нормали к этому элементу, то, интегрируя (15) по площади элемента и используя теорему Стокса, получим

$$\int (\operatorname{rot} n\mathbf{s})\cdot \mathbf{b}\,dS = \int n\,\mathbf{s}\cdot d\mathbf{r} = 0, \tag{16}$$

где второй интеграл берется по ограничивающему элемент контуру $P_1Q_1Q_2P_2$. Переходя к пределу, когда высота $\delta h \to 0$, совершенно таким же способом, как и при выводе (1.1.23), найдем

$$\mathbf{n}_{12}\times(n_2\mathbf{s}_2 - n_1\mathbf{s}_1) = 0, \tag{17}$$

где $\mathbf{n}_{12}$— единичный вектор нормали к поверхности раздела, направленный из первой среды во вторую. Из (17) следует, что *тангенциальная составляющая лучевого вектора $n\mathbf{s}$ непрерывна при переходе через поверхность раздела*, или, что то же самое, *вектор $\mathbf{N} = n_2\mathbf{s}_2 - n_1\mathbf{s}_1$ перпендикулярен к этой поверхности.*

Пусть θ_1 и θ_2— углы, которые образуют падающий и преломленный лучи с нормалью к поверхности $\mathbf{n}_{12}$ (рис. 3.8, *а*). Тогда, согласно (17), имеем

$$n_2(\mathbf{n}_{12}\times\mathbf{s}_2) = n_1(\mathbf{n}_{12}\times\mathbf{s}_1) \tag{18}$$

или

$$n_2\sin\theta_2 = n_1\sin\theta_1. \tag{19}$$

Смысл формулы (18) состоит в том, что *преломленный луч лежит в плоскости, образованной падающим лучом и нормалью к поверхности раздела (плоскости падения)*, а формулы (19) — в том, что *отношение синуса угла преломления к синусу угла падения равно отношению показателей преломления n_1/n_2*. Эти два ре-

зультата выражают *закон преломления* (*закон Снелиуса*). Он был выведен нами ранее в § 1.5 для частного случая плоских волн. Однако если прежний вывод справедлив для случая падения плоской волны с *произвольным* значением λ_0 на плоскую отражающую поверхность, настоящий относится к волнам и отражающим поверхностям более общей формы при условии, что длина волны достаточно мала ($\lambda_0 \to 0$). Последнее условие практически означает, что радиусы кривизны волнового фронта падающей волны и поверхности раздела должны быть велики по сравнению с длиной волны падающего света.

Как и в случае, рассмотренном в § 1.5, следует ожидать, что и здесь появится другая отраженная волна, возвращающаяся обратно в первую среду. Полагая

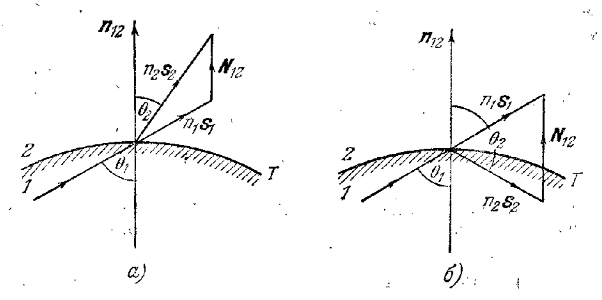


Рис. 3.8. К выводу законов преломления (*а*) и отражения (*б*).

в (18) и (19) $n_2 = n_1$ (см. рис. 3.8, *б*), получим, что *отраженный луч лежит в плоскости падения*, а $\sin\theta_2 = \sin\theta_1$ и, следовательно,

$$\theta_2 = \pi - \theta_1. \tag{20}$$

Последние два результата выражают *закон отражения*.

3.2.3. Конгруэнции лучей и фокальные свойства. Соотношение (15), а именно

$$\operatorname{rot} n\mathbf{s} = 0, \tag{21}$$

определяет все системы лучей, которые могут существовать в изотропной среде, и выделяет их из более общих семейств кривых. В однородной изотропной среде показатель преломления n постоянен, и поэтому (21) принимает вид

$$\operatorname{rot} \mathbf{s} = 0. \tag{22}$$

В неоднородной изотропной среде лучи также можно описать соотношением, не зависящим от n. Его можно получить из (21), если воспользоваться тождеством $\operatorname{rot} n\mathbf{s} = n \operatorname{rot} \mathbf{s} + (\operatorname{grad} n) \times \mathbf{s}$, а затем умножить соотношение (21) скалярно на $\mathbf{s}$. В результате получим, что система лучей в любой *изотропной* среде должна удовлетворять соотношению

$$\mathbf{s} \cdot \operatorname{rot} \mathbf{s} = 0. \tag{23}$$

Система кривых, заполняющих некоторую часть пространства так, что через каждую точку данной области в общем случае проходит одна кривая, называется *конгруэнцией*. Говорят, что конгруэнция *нормальна*, если существует семейство поверхностей, пересекающих каждую кривую под прямым углом; если такого семейства нет, то говорят о *косой* конгруэнции. В обычной геометрической оптике (распространение света) рассматривают только нормальные конгруэнции, однако в электронной оптике (см. приложение 2) важную роль играют и косые конгруэнции.

Если все линии, составляющие конгруэнцию, имеют вид прямых, то такая конгруэнция называется *прямолинейной*; формулы (23) и (24) служат необхо-

димыми и достаточными условиями того, что конгруэнция является соответственно *нормальной* и *нормальной прямолинейной* *).

Выберем на одной из ортогональных поверхностей $\mathcal{S}(x, y, z) = \text{const}$ систему криволинейных координат u, v. Каждой точке $Q(u, v)$ на этой поверхности соответствует одна кривая конгруэнции, а именно кривая, пересекающая данную поверхность в точке Q. Пусть $\mathbf{r}$—радиус-вектор точки P, расположенной на такой кривой. Величину $\mathbf{r}$ можно рассматривать как функцию координат (u, v) и длины дуги s от точки Q до P, измеренной вдоль кривой (рис. 3.9).

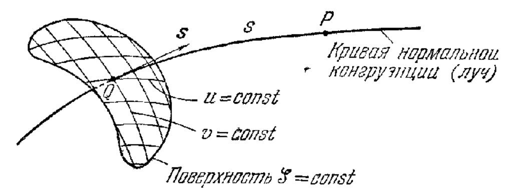


Рис. 3.9. К введению понятия нормальной конгруэнции.

Рассмотрим две соседние кривые из конгруэнции, пересекающие поверхность $\mathcal{S} = \text{const}$ в точках (u, v) и $(u + du, v + dv)$, и выясним, имеются ли на этих кривых такие точки, расстояние между которыми второго или более высокого порядка малости (говорят, что кривые в таких точках имеют пересечение первого порядка). Точки подобного типа называются *фокусами* и удовлетворяют уравнению

$$\mathbf{r}(u, v, s) = \mathbf{r}(u + du, v + dv, s + ds) \tag{24}$$

с точностью до членов второго порядка малости.

Разлагая (24) в ряд, получим

$$\mathbf{r}_u du + \mathbf{r}_v dv + \mathbf{s}\, ds = 0, \tag{25}$$

где $\mathbf{r}_u$, $\mathbf{r}_v$ — частные производные по u и v. Из условия (25) следует, что $\mathbf{r}_u$, $\mathbf{r}_v$ и $\mathbf{s}$ компланарны. Это эквивалентно утверждению, что смешанное произведение этих трех векторов равно нулю, т. е.

$$[\mathbf{r}_u, \mathbf{r}_v, \mathbf{s}] = 0. \tag{26}$$

Число фокусов на данной кривой зависит от числа значений s, удовлетворяющих уравнению (26). Если $\mathbf{r}$ есть полином по s степени m, то, поскольку $\mathbf{s} = d\mathbf{r}/ds$, (26) является уравнением относительно s степени $3m - 1$. В частности, если конгруэнция прямолинейна, то $\mathbf{r}$ — линейная функция $s (m = 1)$, и, следовательно, на *каждом луче прямолинейной конгруэнции имеются два фокуса.*

Если величины u и v принимают все возможные значения, то фокусы образуют поверхность, которая описывается уравнением (26) и называется *фокальной поверхностью*, в оптике она называется *каустической поверхностью*. Любая кривая данной конгруэнции касается фокальной поверхности в каждом своем фокусе. Плоскость, касающаяся фокальной поверхности в какой-либо ее точке, называется *фокальной плоскостью*.

В дальнейшем мы будем в основном изучать поведение лучей в однородной среде, т. е. рассматривать лишь прямолинейные конгруэнции. В § 4.6 при рассмотрении астигматических пучков лучей мы обсудим и некоторые другие свойства таких конгруэнций.

§ 3.3. Другие основные теоремы геометрической оптики

С помощью соотношений, полученных выше, выведем несколько теорем о поведении лучей и волновых фронтов.

3.3.1. Интегральный инвариант Лагранжа. Сначала предположим, что показатель преломления n является непрерывной функцией координат. Тогда,

*) Для более подробного ознакомления с теорией конгруэнций кривых см., например, [27].

если применить теорему Стокса к интегралу от нормальной компоненты rot $n\mathbf{s}$, взятому по любой открытой поверхности, получим (как и при выводе (3.2.16))

$$\oint n\mathbf{s}\cdot d\mathbf{r}=0. \tag{1}$$

Интегрирование здесь проводится по замкнутому контуру C, ограничивающему указанную поверхность. Полученное соотношение называется *интегральным инвариантом Лагранжа* *) и означает, что *интеграл*

$$\int_{P_1}^{P_2} n\mathbf{s}\cdot d\mathbf{r}, \tag{2}$$

взятый между любыми двумя точками поля P_1 и P_2, *не зависит от пути интегрирования.*

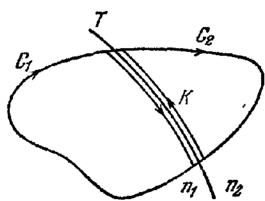


Рис 3 10 К выводу интегрального инварианта Лагранжа при наличии у функции показателя преломления поверхности разрыва.

С помощью закона преломления легко показать, что формула (1) остается справедливой, если контур C пересекает поверхность, разделяющую две однородные среды с разными показателями преломления. Для доказательства положим, что контур C разделяется на части C_1 и C_2, расположенные по разные стороны от преломляющей поверхности T (рис. 3 10), а точки пересечения контура C с поверхностью T соединены другой кривой K, лежащей на этой поверхности. Применяя (1) к обоим контурам C_1K и C_2K и складывая полученные уравнения, имеем

$$\int_{C_1} n_1\mathbf{s}_1\cdot d\mathbf{r}+\int_{C_2} n_2\mathbf{s}_2\cdot d\mathbf{r}+\int_{K}(n_2\mathbf{s}_2-n_1\mathbf{s}_1)\cdot d\mathbf{r}=0. \tag{3}$$

Интеграл вдоль K равен нулю, поскольку вектор $\mathbf{N}=n_1\mathbf{s}_1-n_2\mathbf{s}_2$, согласно закону преломления, нормален поверхности T в любой точке кривой K, и, следовательно выражение (3) сводится к (1).

3.3.2. Принцип Ферма *Принцип Ферма*, известный также как *принцип наикратчайшего оптического пути* **), утверждает, что *оптическая длина*

$$\int_{P_1}^{P_2} n\,ds \tag{4}$$

реального луча между любыми двумя точками P_1 *и* P_2 *короче оптической длины любой другой кривой, соединяющей эти точки и лежащей в некоторой регулярной окрестности луча* Под регулярной окрестностью понимается область, которую можно заполнить лучами таким образом, что через каждую ее точку будет проходить один (и только один) луч. Например, такой областью является та часть пространства, которую заполняют лучи от точечного источника P_1, где эти лучи (из-за преломления, отражения или из-за своей кривизны) не пересекаются.

Прежде чем доказывать это, необходимо отметить, что принцип Ферма можно сформулировать в несколько иной, более слабой форме, применимой, однако, в более широкой области. Согласно данной формулировке реальный

*) Иногда это выражение называют *инвариантом Пуанкаре* Фактически оно является лишь частным одномерным случаем гораздо более общих интегральных инвариантов, рассмотренных Пуанкаре [28, 29] См также приложение 1, уравнение (85).

**) Поскольку из (3 1 27) следует, что

$$\int_{P_1}^{P_2} n\,ds=c\int_{P_1}^{P_2} dt,$$

его называют также *принципом наименьшего времени*.

луч отличается от остальных кривых (не обязательно лежащих в регулярной окрестности) тем, что соответствующий ему интеграл (4) имеет *стационарное значение*.

Чтобы найти кривые, для которых интеграл имеет стационарное значение, необходимо в общем случае применить методы вариационного исчисления, описанные в приложении I. Там показано, что координаты таких кривых удовлетворяют дифференциальным уравнениям Эйлера (см. (7) приложения I). В нашем случае это просто уравнения лучей (3.2.2), что показано в п. 11 приложения I.

Каратеодори [2] подчеркивал, что стационарное значение никогда не является истинным максимумом. Поэтому во второй, более слабой формулировке принципа Ферма необходимо говорить о стационарном, а не экстремальном значении. Исходной же формулировке принципа соответствует «сильный минимум» в смысле Якоби (см. приложение I, п. 10).

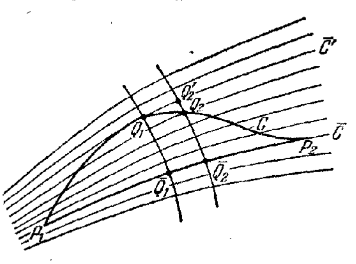


Рис. 3.11. К доказательству принципа Ферма.

Для доказательства принципа Ферма рассмотрим пучок лучей и сравним оптические длины отрезка P_1P_2 луча $\bar{C}$ и произвольной кривой C, соединяющей точки P_1 и P_2 (рис. 3.11). Пусть два соседних волновых фронта пучка пересекают C в точках Q_1 и Q_2, а $\bar{C}$ — в точках $\bar{Q}_1$ и $\bar{Q}_2$. Далее обозначим через Q_2' точку пересечения волнового фронта $Q_2\bar{Q}_2$ с лучом $\bar{C}'$, проходящим через точку Q_1. Применяя интегральное соотношение Лагранжа к маленькому треугольнику $Q_1Q_2Q_2'$, получим

$$(n\mathbf{s}\cdot d\mathbf{r})_{Q_1Q_2}+(n\mathbf{s}\cdot d\mathbf{r})_{Q_2Q_2'}-(n\,ds)_{Q_1Q_2'}=0. \quad (5)$$

Из определения скалярного произведения следует, что

$$(n\mathbf{s}\cdot d\mathbf{r})_{Q_1Q_2}\leqslant(n\,ds)_{Q_1Q_2}.$$

Далее, вектор $\mathbf{s}$ перпендикулярен к $d\mathbf{r}$ на волновом фронте, и поэтому

$$(n\mathbf{s}\cdot d\mathbf{r})_{Q_2Q_2'}=0.$$

Поскольку Q_1, Q_2' и $\bar{Q}_1$, $\bar{Q}_2$ являются соответствующими точками на двух волновых фронтах, находим, согласно (3.1.25),

$$(n\,ds)_{Q_1Q_2'}=(n\,ds)_{\bar{Q}_1\bar{Q}_2}.$$

Применяя в соотношении (5) последние три выражения, получим

$$(n\,ds)_{\bar{Q}_1\bar{Q}_2}\leqslant(n\,ds)_{Q_1Q_2} \quad (6)$$

и после интегрирования —

$$\int_{\bar{C}}n\,ds\leqslant\int_{C}n\,ds. \quad (7)$$

Кроме того, знак равенства можно ставить лишь в том случае, если направления $\mathbf{s}$ и $d\mathbf{r}$ совпадают в каждой точке кривой C, т. е. если она является реальным лучом. Однако такой случай исключен нашим предположением, что через каждую точку окрестности проходит только один луч. Следовательно, оптическая длина луча меньше оптической длины произвольной кривой C, т. е. принцип Ферма доказан.

Легко показать, что в случае невыполнения условия регулярности оптическая длина луча может оказаться не минимальной. Рассмотрим, например, поле лучей от точечного источника P_1 в однородной среде, отраженных плоским зеркалом (рис. 3.12). Через любую точку P_2 в этом случае проходят два луча; оптическая длина прямого луча P_1P_2 является абсолютно минимальной, тогда

как оптическая длина отраженного луча P_1MP_2 минимальна лишь по отношению к оптическим длинам кривых, лежащих в некоторой ограниченной окрестности луча. В общем случае, если лучи от точечного источника P_1 преломляются или отражаются на поверхностях раздела однородных сред, то регулярная область оканчивается на огибающей (каустике) совокупности лучей. Точка P_1', в которой луч от точечного источника P_1 касается огибающей, называется точкой, *сопряженной* точке P_1 на данном луче. Для того чтобы оптическая длина луча P_1P_2 — была минимальной, точка P_2 должна лежать между P_1 и P_1', т. е. точки P_1 и P_2 должны лежать по одну сторону от каустики. Например, в случае нескорректированной линзы (рис. 3.13) центральный луч от P_1 имеет минимальную оптическую длину лишь до кончика каустики P_1' (гауссово изображение точки P_1). Для любой точки P_2, лежащей за огибающей, оптическая длина прямого пути $P_1P_1'P_2$ превышает оптическую длину ломаного пути P_1ABP_2.

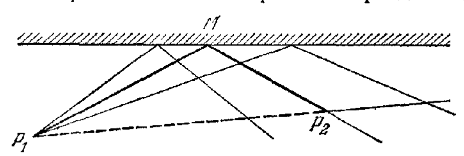


Рис. 3.12. Поле лучей, образующихся при отражении света точечного источника от плоского зеркала.

3.3.3. Теорема Малюса и Дюпина и некоторые другие связанные с ней теоремы. Световые лучи были определены как траектории, ортогональные к волновым поверхностям $\mathcal{S}(x, y, z) = \text{const}$, где $\mathcal{S}$ — решение уравнения эйконала (3.1.15). Такое определение естественно при выводе законов геометрической оптики из уравнений Максвелла. Однако исторически геометрическая оптика развивалась как теория световых лучей, определенных по-другому, а именно как кривых, для которых криволинейные интегралы $\int n\,ds$ имеют стационарные значения. Геометрическую оптику, сформулированную таким образом, можно развивать далее, только используя аппарат вариационного исчисления *).

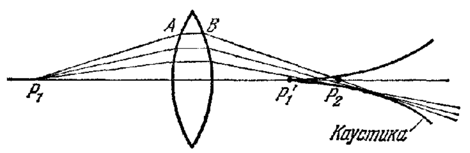


Рис. 3.13. Каустика, образованная лучами от точечного источника, расположенного на оси, после их прохождения через линзу.

Вариационный метод играет в физике очень важную роль, так как он часто раскрывает аналогии между ее различными областями. В частности, существует глубокая аналогия между геометрической оптикой и механикой движущейся частицы, эта аналогия была отчетливо обнаружена благодаря блестящим исследованиям Гамильтона, метод которого приобрел весьма важное значение в современной физике, особенно в волновой механике де Бройля. Чтобы не прерывать рассмотрение оптических проблем, изложение способов вычисления вариаций и описание аналогии Гамильтона помещены в отдельных разделах (см. приложения 1 и 2). Здесь мы только покажем, как из интегрального инварианта Лагранжа можно получить несколько теорем, сыгравших важную роль в развитии геометрической оптики.

Рассмотрим лучи в однородной среде: если все они имеют общую точку, например исходят из точечного источника, то говорят, что лучи образуют *гомоцентрический пучок*. Такой пучок образует нормальную конгруэнцию, поскольку каждый луч пучка пересекает под прямым углом сферические поверхности, центр которых расположен в точке пересечения лучей.

*) Систематическое изложение такого типа приведено, например, в работе Каратеодори [2].

Малюс *) в 1808 г. показал, что если гомоцентрический пучок прямолинейных лучей преломляется или отражается какой-нибудь поверхностью, то получающийся после этого пучок (в общем случае уже не гомоцентрический) тоже образует нормальную конгруэнцию. Позднее Дюпин (1816 г.), Кветеле (1825 г.) и Жергонн (1825 г.) обобщили результат Малюса. Работы этих ученых позволили сформулировать следующую теорему, называемую иногда *теоремой Малюса и Дюпина*. *Нормальная прямолинейная конгруэнция остается нормальной после любого числа преломлений и отражений* **).

Достаточно доказать эту теорему для случая одного акта преломления. Рассмотрим нормальную прямолинейную конгруэнцию лучей в однородной среде с показателем преломления n_1 и предположим, что лучи преломляются на поверхности T, отделяющей эту среду от другой однородной среды с показателем преломления n_2 (рис. 3.14).

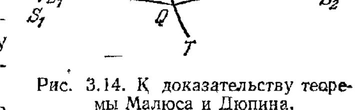


Рис. 3.14. К доказательству теоремы Малюса и Дюпина.

Пусть S_1 — один из волновых фронтов в первой области, A_1 и P — точки пересечения произвольного луча в первой среде соответственно с S_1 и T, а A_2 — точка на преломленном луче. Если точку A_1 сместить в другую точку B_1 на том же волновом фронте, то точка P на поверхности преломления сместится в Q. Теперь на преломленном в точке Q луче выберем такую точку B_2, чтобы оптический путь от B_1 до B_2 равнялся оптическому пути от A_1 до A_2, т. е. чтобы

$$[A_1PA_2] = [B_1QB_2]. \tag{8}$$

Если перемещать точку B_1 по всей поверхности S_1, то точка B_2 при своем перемещении заполнит поверхность S_2. Покажем, что преломленный луч QB_2 перпендикулярен к этой поверхности.

Вычисляя интегральный инвариант Лагранжа по замкнутому пути $A_1PA_2B_2QB_1A_1$, получим

$$\int\limits_{A_1PA_2} n\,ds + \int\limits_{A_2B_2} n\mathbf{s}\cdot d\mathbf{r} + \int\limits_{B_2QB_1} n\,ds + \int\limits_{B_1A_1} n\mathbf{s}\cdot d\mathbf{r} = 0. \tag{9}$$

На основании (8) можно написать

$$\int\limits_{A_1PA_2} n\,ds + \int\limits_{B_2QB_1} n\,ds = 0. \tag{10}$$

Кроме того, поскольку на волновом фронте S_1 единичный вектор $\mathbf{s}$ всюду перпендикулярен к нему, имеем

$$\int\limits_{B_1A_1} n\mathbf{s}\cdot d\mathbf{r} = 0 \tag{11}$$

и, следовательно, (9) принимает вид

$$\int\limits_{A_2B_2} n\mathbf{s}\cdot d\mathbf{r} = 0. \tag{12}$$

Полученное соотношение должно выполняться на S_2 для любого отрезка кривой. Это возможно только в том случае, если $\mathbf{s}\cdot d\mathbf{r} = 0$ для каждого линей-

*) См. [30], а также [31]. В работе [32] приведены ссылки и изложена интересная история теоремы Малюса — Дюпина.

**) Леви Чивита [33] доказал обратную теорему, состоящую в том, что две любые нормальные прямолинейные конгруэнции можно перевести друг в друга с помощью одного преломления или отражения.

ного элемента $d\mathbf{r}$ поверхности S_2, т. е. если преломленные лучи перпендикулярны к ней, другими словами, если *преломленные лучи образуют нормальную конгруэнцию*. Доказательство для случая отражения абсолютно аналогично приведенному выше.

Поскольку $[A_1PA_2]=[B_1QB_2]$, можно утверждать, что оптическая *длина пути между любыми двумя волновыми фронтами одинакова для всех лучей.* Очевидно, что этот результат остается справедливым для случая нескольких последовательных преломлений или отражений, а также, как непосредственно следует из (3.1.26), в случае распространения лучей в среде с непрерывно изменяющимся показателем преломления. Эта теорема называется *принципом равного оптического пути*, из нее следует, что геометрические волновые фронты нормальной конгруэнции лучей или совокупности нормальных конгруэнций, образованных в результате последовательных преломлении или отражений, «оптически параллельны» друг другу (см. приложение I).

С последней теоремой связана теорема, впервые выдвинутая Гюйгенсом [34], которая утверждает, что *каждый элемент волнового фронта можно рассматривать как центр вторичного возмущения, порождающего вторичные сферические волны* и, кроме того, что *волновым фронтом в любой последующий момент времени служит огибающая этих вторичных сферических волн*. Это утверждение (*построение Гюйгенса*) служит правилом для построения поверхностей, «оптически параллельных» друг другу. Если среда однородна, то при построении можно использовать сферические волны конечного радиуса, в противном случае необходимо пользоваться волнами бесконечно малого радиуса.

Теорема Гюйгенса была позднее обобщена Френелем и легла в основу так называемого принципа Гюйгенса — Френеля, играющего важную роль в теории дифракции (см. § 8.2) и являющегося основным постулатом волновой теории света.

ЛИТЕРАТУРА

1. M. Herzberger, Strahlenoptik, Springer, Berlin, 1931, p. 179, Z. f. Instrkde 52, 429—435, 485—493, 534—542 (1932).
2. C. Caratheodory, Geometrische Optik, Springer, Berlin, 1937.
3. E. Mach, The Principles of Physical Optics, A Historical and Philosophical Treatment (первое немецкое издание 1913 г.)
4. G. Kirchhoff, Vorlesungen u. Math. Phys. 2 (Mathematische Optik), Teubner, Leipzig, 1891, p. 33.
5. B. Baker, E. Copson, The Mathematical Theory of Huygens' Principle, Clarendon Press, Oxford, 2nd ed. 1950, p. 79.
6. A. Sommerfeld, Optics, Acad. Press, New York, 1954. (А. Зоммерфельд, Оптика, ИЛ, 1953.)
7. A. Sommerfeld, J. Runge, Ann. d. Physik 35, 289 (1911).
8. В. Игнатовский, Труды Оптического института, Петроград, I, III (1919).
9. В. А. Фок, Труды Оптического института 3, 3 (1924).
10. С. М. Рытов, ДАН СССР 18, 263 (1938).
11. N. Arley, Det Kgl. Danske Vidensk. Selsk. 22, № 8, 1945.
12. F. G. Friedlander, Proc. Cambr. Phil. Soc. 43, 284 (1947).
13. K. Suchy, Ann. d. Physik 11, 113 (1952), 12, 423 (1953), 13, 178 (1953).
14. R. S. Ingarden, A. Krzywicki, Acta Phys. Polonica 14, 255 (1955).
15. R. K. Luneburg, Mathematical Theory of Optics (Мимеограф), Brown Univ., Providence, 1944, p. 55—59, печатное издание Univ. California Press, Berkeley a. Los Angeles, 1964, p. 51—55.
16. R. K. Luneburg, Propagation of Electromagnetic Waves, New York, University, 1947—1948.
17. M. Kline, I. W. Kay, Electromagnetic Theory of Geometrical Optics, Intersci. Publ., New York, 1965.

17a. M. Kline, Comm. Pure a. Appl. Math., 8, 595 (1955).

18. M. Kline, Comm. Pure a. Appl. Math. 4, 225 (1951).
19. Symposium on Theory of Electromagnetic Waves, Interscience New York, 1951, p. 225.
20. M. Kline, Comm. Pure a. Appl. Math. 14, 473 (1961).
21. Г. Bortolotti, Rend. R. Acc. Naz. Linc. 6a 4, 552 (1926).
22. W. Braunbek, Z. Naturforsch. 6, 672 (1951).

23. J. B. Keller, J. Appl. Phys. **28**, 426 (1957); см. также Calculus of Variations and its Applications, ed. L. M. Graves, McGrow-Hill, New York, 1958.
24. J. B. Keller, Trans. IRE, A. P.-4, 312 (1956).
25. J. B. Keller, R. M. Lewis, B. D. Seckler, J. Appl. Phys. **28**, 570 (1957).
26. R. Courant, Differential and Integral Calculus, Vol. I., Blackie, Glasgow, 2-nd ed., 1942 (Р. Курант, Курс дифференциального и интегрального исчисления, ГТТИ, 1931.)
27. C. E. Weatherburn, Differential Geometry of Three Dimensions, Cambr. Univ. Press., Vol. I, 1927; Vol. II, 1930.
28. J. H. Poincaré, Les Méthodes Nouvelles de la Mécanique Céleste, Gauthier-Villars, Paris, 1899, Vol. 3.
29. E. Cartan, Leçons sur les Invariants Integraux, Hermann, Paris, 1922. (Э. Картан, Интегральные инварианты, Гостехиздат, 1940.)
30. E. Malus, Optique Dioptrique, J. Ecole polytechn. **7**, 1—44, 84—129 (1808).
31. E. Malus, Traité d'Optique, Mém. présent à l'Institut par divers savants **2**, 214—302 (1811).
32. W. R. Hamilton, Mathematical Papers, Vol. 1, Geometrical Optics, ed. by A. W. Conway, a. J. L. Synge, Cambr. Univ. Press. 1931, p. 463.
33. T. Levi-Civita, Rend. R. Acc. Naz. Linc. **9**, 237 (1900).
34. Chr. Huygens, Traité de la Lumiére, Leiden, 1690; MacMillan a. Co. 1912. (Х. Гюйгенс, Трактат о свете, ОНТИ, 1935.)

ГЛАВА 4

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ

§ 4.1. Характеристические функции Гамильтона

В § 3.1 было показано, что в приближении геометрической оптики поле можно охарактеризовать одной скалярной функцией $\mathcal{S}(\mathbf{r})$. Поскольку $\mathcal{S}(\mathbf{r})$ удовлетворяет уравнению эйконала (3.1.15), эта функция полностью определяется величиной показателя преломления $n(\mathbf{r})$ и соответствующими граничными условиями.

Часто вместо функции $\mathcal{S}(\mathbf{r})$ используют тесно связанные с ней функции, называемые *характеристическими функциями* среды. Они были введены в оптику Гамильтоном в серии его классических статей [1] *). Несмотря на то, что вследствие существенных алгебраических трудностей характеристические функции можно вычислить в явном виде лишь для самых простых сред, методы Гамильтона являются мощным средством для систематических аналитических исследований общих свойств оптических систем.

Рис. 4.1. К определению точечной характеристики.

При обсуждении свойств этих функций и их приложений мы будем предполагать, что рассматриваемая среда изотропна, но, вообще говоря, неоднородна.

4.1.1. Точечная характеристика. Пусть (x_0, y_0, z_0) и (x_1, y_1, z_1) — соответственно координаты двух точек P_0 и P_1 в двух различных прямоугольных координатных системах, оси которых параллельны друг другу **) (рис. 4.1). Если соединить эти точки всеми возможными кривыми, то в общем случае некоторые из них окажутся оптическими лучами, удовлетворяющими принципу Ферма. Предположим вначале, что две произвольные точки соединяются только одним лучом. Тогда характеристическая функция V, или *точечная характеристика*, определяется как *оптическая длина* $[P_0P_1]$ *луча между двумя точками*, рассматриваемая как функция их координат, т. е.

$$V(x_0, y_0, z_0; x_1, y_1, z_1) = \int_{P_0}^{P_1} n\, ds. \quad (1)$$

Необходимо подчеркнуть, что эта функция определяется *свойствами среды*.

*) Эти работы перепечатаны в [2]. Много лет спустя Брунс независимо рассмотрел аналогичные функции, которые он назвал *эйконалами* [3]. Как уже отмечалось выше, этот термин стал употребляться позднее в более широком смысле. Сами характеристические функции Гамильтона тоже иногда называют эйконалами.

Полезным введением к изучению методов Гамильтона служит монография Синджа [4]. Связь между работами Гамильтона и Брунса обсуждалась в трудах Клейна [5, 6], Каратеодори [7] и в полемике Герцбергера с Синджем [8].

**) Использование двух координатных систем имеет некоторое преимущество, поскольку точки P_0 и P_1 часто расположены в различных областях, а именно в пространстве объекта и пространстве изображения оптической системы.

Из выражения (1) и соотношения (3.1.26) следует, что

$$V(x_0, y_0, z_0; x_1, y_1, z_1) = \mathcal{S}(x_1, y_1, z_1) - \mathcal{S}(x_0, y_0, z_0), \tag{2}$$

где функция $\mathcal{S}$ связана с любым пучком лучей, к которому принадлежит световой луч, соединяющий точки P_0 и P_1 (например, с пучком, испускаемым точечным источником в P_0) *). Тогда на основании (3.1.24) получим следующие соотношения для направленных вдоль луча единичных векторов $\mathbf{s}_0$ и $\mathbf{s}_1$ в точках P_0 и P_1:

$$\operatorname{grad}^0 V = -n_0 \mathbf{s}_0, \qquad \operatorname{grad}^1 V = n_1 \mathbf{s}_1, \tag{3}$$

здесь индексы 0 и 1 указывают, что оператор grad действует соответственно на координаты (x_0, y_0, z_0) и (x_1, y_1, z_1).

Вектор

$$\mathbf{g} = n\mathbf{s} \tag{4}$$

иногда называют *лучевым вектором*. Пусть α, β и γ — углы, образованные лучевым вектором с координатными осями; тогда его проекции на оси

$$p = n\cos\alpha, \qquad q = n\cos\beta, \qquad m = n\cos\gamma \tag{5}$$

называются *лучевыми компонентами* **). Учитывая тождество

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1,$$

мы видим, что они удовлетворяют следующему соотношению:

$$p^2 + q^2 + m^2 = n^2. \tag{6}$$

Согласно (3) лучевые компоненты в точках P_0 и P_1 определяются выражениями

$$p_0 = -\frac{\partial V}{\partial x_0}, \qquad p_1 = \frac{\partial V}{\partial x_1}; \tag{7}$$

аналогичные соотношения справедливы для q_0, q_1 и m_0, m_1. Отсюда следует, что, зная точечную характеристику, можно сразу же определить компоненты луча, соединяющего две произвольные точки в среде. Далее из (6) и (7) вытекает, что точечная характеристика удовлетворяет уравнению эйконала, записанному как в координатах x_0, y_0, z_0, так и в координатах x_1, y_1, z_1, т. е.

$$\left(\frac{\partial V}{\partial x_0}\right)^2 + \left(\frac{\partial V}{\partial y_0}\right)^2 + \left(\frac{\partial V}{\partial z_0}\right)^2 = n_0^2, \tag{8}$$

$$\left(\frac{\partial V}{\partial x_1}\right)^2 + \left(\frac{\partial V}{\partial y_1}\right)^2 + \left(\frac{\partial V}{\partial z_1}\right)^2 = n_1^2. \tag{9}$$

Часто вместо точечной характеристики удобно использовать другие, связанные с ней функции (также введенные Гамильтоном), которые называются *смешанной* и *угловой характеристиками*. Их можно получить из точечной характеристики с помощью преобразований Лежандра***), и они оказываются особенно полезными, когда P_0 или P_1, или обе эти точки находятся в бесконечности.

---

*) На языке вариационного исчисления $\mathcal{S}$ служит решением уравнения Гамильтона — Якоби, связанного с вариационной задачей Ферма, содержащим двухпараметрическое семейство (∞^2) экстремалей. Точечная характеристическая функция V является общим решением, содержащим все (∞^4) экстремали (см. приложение I).

**) Такое «асимметричное» обозначение выбрано здесь с определенной целью — напомнить, что благодаря тождеству (6) независимы только две лучевые компоненты.

***) Преобразование Лежандра в общем случае так трансформирует функцию $f(x, y)$ в функцию $g(x, z)$, где $z = \partial f/\partial y$, что производная от g по новой переменной z равна старой переменной y.

4.1.2. Смешанная характеристика. Смешанная характеристическая функция W определяется из уравнения

$$W = V - \sum p_1 x_1, \tag{10}$$

где суммируются три сходных члена с индексом 1.

Чтобы придать смешанной характеристике более наглядный физический смысл, мы, следуя Синджу, определяем ее со знаком, противоположным тому, который использовал Гамильтон

Смешанную характеристику можно также определить соотношением $W' = V + \sum p_0 x_0$, где суммирование производится по сходным членам с индексом 0 Функции W и W' обладают, конечно, совершенно одинаковыми свойствами

Как мы видим, W — функция девяти переменных, однако в общем случае только шесть из них независимы (в однородной среде — пять). Для доказательства последнего утверждения предположим, что точки P_0 и P_1 сместились на небольшие расстояния. Соответствующее изменение W дается выражением

$$\delta W = \delta V - \sum p_1 \delta x_1 - \sum x_1 \delta p_1. \tag{11}$$

Согласно (7)

$$\delta V = \sum p_1 \delta x_1 - \sum p_0 \delta x_0. \tag{12}$$

Из (11) и (12) находим

$$\delta W = -\sum p_0 \delta x_0 - \sum x_1 \delta p_1. \tag{13}$$

Отсюда следует, что в общем случае W является функцией шести переменных $x_0, y_0, z_0, p_1, q_1, m_1$; при этом

$$p_0 = -\frac{\partial W}{\partial x_0}, \quad x_1 = -\frac{\partial W}{\partial p_1}; \tag{14}$$

аналогичные выражения справедливы для q_0, y_1, m_0 и z_1. Согласно (6) функция $W(x_0, y_0, z_0, p_1, q_1, m_1)$ удовлетворяет уравнению эйконала

$$\left(\frac{\partial W}{\partial x_0}\right)^2 + \left(\frac{\partial W}{\partial y_0}\right)^2 + \left(\frac{\partial W}{\partial z_0}\right)^2 = n_0^2. \tag{15}$$

Отметим (рис. 4.2), что сумма $\sum p_1 x_1$ имеет простое геометрическое истолкование, а именно

$$\sum p_1 x_1 = n_1 d_1, \tag{16}$$

где $d_1 = Q_1 P_1$ — проекция $O_1 P_1$ на касательную к лучу в точке P_1 Если P_1 находится в однородной среде, то часть луча в окрестности P_1 совпадает с отрезком $Q_1 P_1$, тогда, согласно (10) и (16), функция W *является оптической длиной луча от точки P_0 до основания Q_1 перпендикуляра, опущенного из начала координат O_1 на выходящий из системы луч* (рис 4 3), т е.

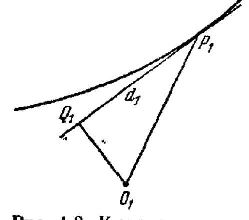


Рис 4 2 К определению смешанной характеристики.

$$W = [P_0 Q_1]. \tag{17}$$

Поскольку в этом случае показатель преломления среды около точки P_1 постоянен, из выражения (6) следует, что

$$\delta m_1 = -\frac{p_1 \delta p_1 + q_1 \delta q_1}{m_1}, \tag{18}$$

а формула (13) после подстановки в нее (18) принимает вид *)

$$\delta W = -\sum p_0 \delta x_0 - \left(x_1 - \frac{p_1}{m_1} z_1\right) \delta p_1 - \left(y_1 - \frac{q_1}{m_1} z_1\right) \delta q_1. \tag{19}$$

*) Если какая либо функция зависит от переменных, связанных дополнительными соотношениями, такими, как (6), то часть этих переменных можно исключить можно также воспользоваться указанными соотношениями для преобразования функции в однородную по всем переменным Последнюю процедуру, которая оказывается довольно сложной, часто использовал Гамильтон.

Следовательно, *если у конца луча среда однородна, то смешанная характеристика является функцией пяти переменных*

$$W = W(x_0, y_0, z_0; p_1, q_1), \tag{20}$$

а ее производные удовлетворяют следующим соотношениям

$$p_0 = -\frac{\partial W}{\partial x_0}, \quad q_0 = -\frac{\partial W}{\partial y_0}, \quad m_0 = -\frac{\partial W}{\partial z_0}, \tag{21}$$

$$x_1 - \frac{p_1}{m_1} z_1 = -\frac{\partial W}{\partial p_1}, \quad y_1 - \frac{q_1}{m_1} z_1 = -\frac{\partial W}{\partial q_1}. \tag{22}$$

Отсюда вытекает, что если заданы точка на луче в первой среде и лучевые компоненты во второй среде, то, зная смешанную характеристику, можно сразу

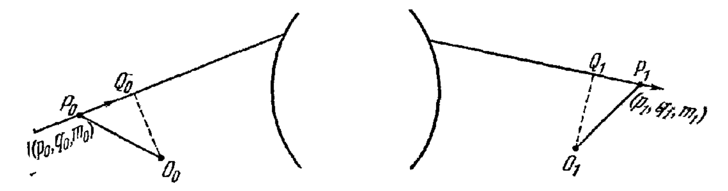


Рис 4 3. Интерпретация характеристических функций Гамильтона в случае однородных начальной и конечной сред

$[V(x_0, y_0, z_0, x_1, y_1, z_1) = [P_0P_1], \; W(x_0, y_0, z_0\; p_1, q_1) = [P_0Q_1], \; T(p_0, q_0\; p_1, q_1) = [Q_0Q_1]$

же определить лучевые компоненты в первой среде и точки на луче во второй среде.

4.1.3. **Угловая характеристика.** Угловую характеристику T можно определить с помощью следующего соотношения:

$$T = V + \sum p_0 x_0 - \sum p_1 x_1. \tag{23}$$

При небольшом смещении точек P_0 и P_1 соответствующее изменение T, согласно (12), равно

$$\delta T = \sum x_0 \delta p_0 - \sum x_1 \delta p_1. \tag{24}$$

Следовательно, T является функцией шести лучевых компонент, и при этом

$$x_0 = \frac{\partial T}{\partial p_0}, \qquad x_1 = -\frac{\partial T}{\partial p_1}; \tag{25}$$

аналогичные выражения справедливы и для других координат

Из (23) следует, что *если среды, в которых находятся точки P_0 и P_1, однородны, то функция T является оптической длиной луча между основаниями Q_0 и Q_1 перпендикуляров, опущенных из O_0 и O_1 на начальный и конечный отрезки луча* (см. рис. 4 3), т. е.

$$T = [Q_0 Q_1]. \tag{26}$$

В этом случае угловую характеристику можно представить как функцию только четырех переменных. В самом деле, если использовать формулу (18) для δm_1 и аналогичную формулу для δm_0, то (24) примет вид

$$\delta T = \left(x_0 - \frac{p_0}{m_0} z_0\right) \delta p_0 + \left(y_0 - \frac{q_0}{m_0} z_0\right) \delta q_0 - \\ - \left(x_1 - \frac{p_1}{m_1} z_1\right) \delta p_1 - \left(y_1 - \frac{q_1}{m_1} z_1\right) \delta q_1. \tag{27}$$

Последнее выражение показывает, что в случае *однородных начальной и конечной сред угловая характеристика зависит от четырех переменных* p_0, q_0, p_1 и q_1,

а именно

$$T = T(p_0,\ q_0;\ p_1,\ q_1), \tag{28}$$

а ее производные удовлетворяют соотношениям

$$\left.\begin{aligned} x_0 - \frac{p_0}{m_0} z_0 &= \frac{\partial T}{\partial p_0}, & y_0 - \frac{q_0}{m_0} z_0 &= \frac{\partial T}{\partial q_0}, \\ x_1 - \frac{p_1}{m} z_1 &= -\frac{\partial T}{\partial p_1}, & y_1 - \frac{q_1}{m_1} z_1 &= -\frac{\partial T}{\partial q_1} \end{aligned}\right\} \tag{29}$$

Итак, если известны лучевые компоненты начального и конечного отрезков луча, то, зная угловую характеристику, можно сразу же с помощью (29) определить координаты точек на этих участках луча

4.1.4. Приближенное выражение для угловой характеристики преломляющей поверхности вращения. Пусть

$$z = c_2(x^2 + y^2) + c_4(x^2 + y^2)^2 - \ldots, \tag{30}$$

где c_2, c_4, ... — постоянные, является уравнением преломляющей поверхности вращения, отнесенным к декартовой системе координат, начало которой O совпадает с осевой точкой (называемой полюсом) поверхности, а ось z направлена вдоль оси симметрии Если через r обозначить радиус кривизны поверхности у полюса (считающийся положительным, когда поверхность обращена выпуклостью к свету, распространяющемуся вдоль положительного направления оси z), то

$$c_2 = 1/2r \tag{31}$$

Для сферической поверхности радиуса r имеем c_4 $1/8r^3$. В случае произвольной поверхности вращения можно написать

$$c_4 - \frac{1}{8r^3}(1 + b), \tag{32}$$

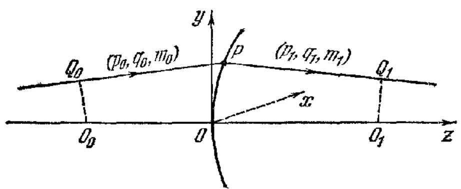


Рис. 4 4 Угловая характеристика преломляющей поверхности вращения

Точки O_0, O; O_1, Q_1, P Q_0 не обязательно компланарны

где константа b (иногда называемая *коэффициентом деформации*) служит грубой мерой отклонения формы поверхности от сферической Уравнение (30), выраженное через r и b, принимает вид

$$z = \frac{x^2 + y^2}{2r} + \frac{(x^2 + y^2)^2}{8r^3}(1 + b) + \ldots \tag{33}$$

Пусть по обе стороны от поверхности вращения среды однородны и обладают показателями преломления n_0 и n_1 Отнесем угловую характеристику к координатным системам, оси которых параллельны осям системы с началом в точке O, а начала расположены в осевых точках $O_0(0,\ 0,\ a_0)$ и $O_1(0,\ 0,\ a_1)$ $(a_0 < 0,\ a_1 > 0,\ r > 0$ на рис 4 4). Если P — точка пересечения падающего луча с преломляющей поверхностью, а Q_0 и Q_1 — основания перпендикуляров, опущенных из O_0 и O_1 на падающий и преломленный лучи, то, согласно (26), угловая характеристика T равна

$$T = [Q_0P] + [PQ_1] = \{xp_0 + yq_0 + (z - a_0)m_0\} - \{xp_1 + yq_1 + (z - a_1)m_1\}, \tag{34}$$

где $(x,\ y,\ z)$ — координаты точки P в системе координат с началом в O, а $(p_0,\ q_0,\ m_0)$, $(p_1,\ q_1,\ m_1)$ — лучевые компоненты падающего и преломленного в точке P лучей.

Координаты $(x,\ y,\ z)$ можно исключить из (34) с помощью закона преломления. Закон преломления, согласно п. 3.2.2, эквивалентен утверждению, что

вектор $\mathbf{N}(p_0-p_1,\ q_0-q_1,\ m_0-m_1)$ нормален к преломляющей поверхности в точке P. Следовательно, если (33) записать в виде

$$F(x, y, z) \equiv z - \frac{x^2+y^2}{2r} + \frac{(x^2+y^2)^2}{8r^3}(1+b) + \ldots = 0, \tag{35}$$

то

$$\left.\begin{aligned} \lambda\frac{\partial F}{\partial x} &= -\lambda\left[\frac{x}{r}+\ldots\right] = p_0-p_1,\\ \lambda\frac{\partial F}{\partial y} &= -\lambda\left[\frac{y}{r}+\ldots\right] = q_0-q_1,\\ \lambda\frac{\partial F}{\partial z} &= \lambda \qquad\qquad\quad = m_0-m_1. \end{aligned}\right\} \tag{36}$$

Из этих уравнений следует, что

$$x = -r\frac{p_0-p_1}{m_0-m_1} + \Delta x, \qquad y = -r\frac{q_0-q_1}{m_0-m_1} + \Delta y, \tag{37}$$

где Δx и Δy — величины третьего порядка по p, q, x/r, y/r. Подставляя (37) в (35), получим формулу для z, выраженную через лучевые компоненты:

$$z = \frac{r}{2(m_1-m_0)^2}\left[(p_0-p_1)^2+(q_0-q_1)^2\right] + \frac{1}{m_1-m_0}\left[\Delta x(p_0-p_1) + \right.$$
$$\left. + \Delta y(q_0-q_1)\right] + \frac{1}{8}\frac{r(1+b)}{(m_1-m_0)^4}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2 + \ldots \tag{38}$$

Для нахождения разложения T с точностью до членов четвертого порядка включительно нет необходимости вычислять величины Δx и Δy, так как после подстановки (37) и (38) в (34) члены, содержащие Δx и Δy, оказываются более высокого, чем четвертый, порядка малости, и поэтому ими можно пренебречь. В результате (34) принимает вид

$$T(p_0, q_0, m_0;\ p_1, q_1, m_1) = -m_0a_0 + m_1a_1 +$$
$$+ \frac{r}{2(m_1-m_0)}\left[(p_0-p_1)^2+(q_0-q_1)^2\right] - \frac{1}{8}\frac{r(1+b)}{(m_1-m_0)^2}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2. \tag{39}$$

Выражение (39) является разложением угловой характеристики до членов четвертого порядка включительно, причем эта характеристика рассматривается как функция всех шести лучевых компонент. Две компоненты можно исключить с помощью равенства (6). Из (6) найдем

$$\left.\begin{aligned} m_0 &= n_0 - \frac{1}{2n_0}(p_0^2+q_0^2) - \frac{1}{8n_0^3}(p_0^2+q_0^2)^2 + \ldots,\\ m_1 &= n_1 - \frac{1}{2n_1}(p_1^2+q_1^2) - \frac{1}{8n_1^3}(p_1^2+q_1^2)^2 + \ldots \end{aligned}\right\} \tag{40}$$

и, следовательно,

$$\frac{1}{m_1-m_0} = \frac{1}{n_1-n_0}\left[1 - \frac{1}{2n_0(n_1-n_0)}(p_0^2+q_0^2) + \frac{1}{2n_1(n_1-n_0)}(p_1^2+q_1^2) + \ldots\right]. \tag{41}$$

Подставляя это в (39), получим

$$T(p_0, q\ ;\ p_1, q_1) = n_1a_1 - n_0a_0 + \frac{r}{2(n_1-n_0)}\left[(p_0-p_1)^2+(q_0-q_1)^2\right] +$$
$$+ \frac{a_0}{2n_0}(p_0^2+q_0^2) - \frac{a_1}{2n_1}(p_1^2+q_1^2) -$$
$$- \frac{r}{4(n_1-n_0)^2}\left[(p_0-p_1)^2+(q_0q_1)^2\right]\left[\frac{p_0^2+q_0^2}{n_0} - \frac{p_1^2+q_1^2}{n_1}\right] -$$
$$- \frac{(1+b)\,r}{8(n_1-n_0)^3}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2 + \frac{a_0}{8n_0^3}(p_0^2+q_0^2)^2 - \frac{a_1}{8n_1^3}(p_1^2+q_1^2)^2 + \ldots \tag{42}$$

Как мы видим, четыре переменные p_0, q_0, p_1 и q_1 входят в это выражение только в комбинациях

$$p_0^2+q_0^2=u^2, \qquad p_1^2+q_1^2=v^2 \quad \text{и} \quad p_0p_1+q_0q_1=w^2. \tag{43}$$

Можно показать в более общем виде, что угловая характеристика любой среды, вращательно симметричной относительно оси z, зависит от четырех переменных, входящих только в трех комбинациях (43). Для доказательства этого используем результат, полученный ниже, в § 5.1, согласно которому любая функция $F(x_0, y_0; x_1, y_1)$, инвариантная относительно вращения осей вокруг начала координат в плоскости xy, зависит только от величин трех скалярных произведений

$$\mathbf{r}_0^2=x_0^2+y_0^2, \qquad \mathbf{r}_1^2=x_1^2+y_1^2, \qquad \mathbf{r}_0\cdot\mathbf{r}_1=x_0x_1+y_0y_1$$

двух векторов $\mathbf{r}_0(x_0, y_0)$ и $\mathbf{r}_1(x_1, y_1)$. Считая $\mathbf{r}_0$ и $\mathbf{r}_1$ проекциями векторов распространения $\mathbf{g}_0(p_0, q_0, m_0)$ и $\mathbf{g}_1(p_1, q_1, m_1)$ на плоскости xy, получим требуемый результат.

Подставляя (43) в (42) и выделяя члены одного порядка, получим окончательно

$$T(p_0, q_0; p_1, q_1)=T^{(0)}+T^{(2)}+T^{(4)}+\ldots,$$

где

$$\left.\begin{aligned} T^{(0)}&=n_1a_1-n_0a_0,\\ T^{(2)}&=\mathcal{A}u^2+\mathcal{B}v^2+\mathcal{C}w^2,\\ T^{(4)}&=\mathcal{D}u^4+\mathcal{E}v^4+\mathcal{F}w^4+\mathcal{G}u^2v^2+\mathcal{H}u^2w^2+\mathcal{K}v^2w^2 \end{aligned}\right\} \tag{44}$$

и

$$\left.\begin{aligned} \mathcal{A}&=\frac{1}{2}\left[\frac{r}{n_1-n_0}+\frac{a_0}{n_0}\right],\\ \mathcal{B}&=\frac{1}{2}\left[\frac{r}{n_1-n_0}-\frac{a_1}{n_1}\right],\\ \mathcal{C}&=-\frac{r}{n_1-n_0},\\ \mathcal{D}&=-\frac{r}{4(n_1-n_0)^2}\left[\frac{1+b}{2(n_1-n_0)}+\frac{1}{n_0}\right]+\frac{a_0}{8n_0^3},\\ \mathcal{E}&=-\frac{r}{4(n_1-n_0)^2}\left[\frac{1+b}{2(n_1-n_0)}-\frac{1}{n_1}\right]-\frac{a_1}{8n_1^3},\\ \mathcal{F}&=\frac{-(1+b)r}{2(n_1-n_0)^3},\\ \mathcal{G}&=-\frac{r}{4(n_1-n_0)^2}\left[\frac{1+b}{n_1-n_0}+\frac{1}{n_0}-\frac{1}{n_1}\right],\\ \mathcal{H}&=\frac{r}{2(n_1-n_0)^2}\left[\frac{1+b}{n_1-n_0}+\frac{1}{n_0}\right],\\ \mathcal{K}&=\frac{r}{2(n_1-n_0)^2}\left[\frac{1+b}{n_1-n_0}-\frac{1}{n_1}\right]. \end{aligned}\right\} \tag{45}$$

4. 1.5. Приближенное выражение для угловой характеристики отражающей поверхности вращения. Разложение угловой характеристики отражающей поверхности вращения до членов четвертого порядка включительно можно провести аналогичным способом. Однако нет необходимости повторять полностью все вычисления. Если использовать те же обозначения, что и выше (ср. рис. 4.4 и 4.5), то все уравнения из п. 4.1.4 до соотношения (39) включительно остаются прежними; следовательно, для *отражающей поверхности вращения* (39) является также *угловой характеристикой, рассматриваемой как функция всех шести лучевых компонент.* Однако при исключении m_0 и m_1 из (39) с помощью двух соотношений, связывающих лучевые компоненты, для T (как функции четырех лучевых компонент) в случаях преломляющей и отражающей поверхностей получаются два различных выражения. Если обозначить показатель преломления

среды, в которой распространяются лучи, через n, то вместо (40) найдем

$$\left.\begin{aligned} m_0 &= n - \frac{1}{2n}(p_0^2 + q_0^2) - \frac{1}{8n^3}(p_0^2 + q_0^2)^2 + \ldots, \\ m_1 &= -\left[n - \frac{1}{2n}(p_1^2 + q_1^2) - \frac{1}{8n^3}(p_1^2 + q_1^2)^2 + \ldots\right]. \end{aligned}\right\} \quad (46)$$

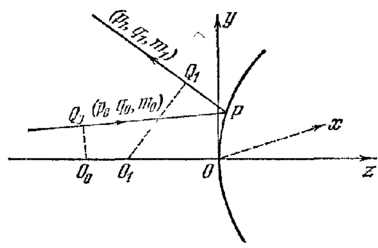


Рис 4 5 Угловая характеристика отражающей поверхности вращения Точки O_0 O_1, O Q_0, P, Q_1 не обязательно компланарны

Для второго из этих соотношении выбран отрицательный корень $-\sqrt{n^2-(p^2+q^2)}$, поскольку отраженный луч попадает обратно в ту область, из которой он приходит ($z<0$), поэтому направляющий косинус отраженного луча относительно положительного направления z, а следовательно, и m отрицательны Поскольку (40) переходит в (46) при замене $n_0 = -n_1 = n$, можно заключить, что *угловая характеристика отражающей поверхности вращения, рассматриваемая как функция четырех лучевых компонент p_0, q_0, p_1 и q_1, получается из угловой характеристики преломляющей поверхности вращения, если положить $n_0 = -n_1 = n$* Следовательно, для случая отражения имеем

$$\left.\begin{aligned} T^{(0)} &= -n(a_0 + a_1), \\ T^{(2)} &= \mathcal{A}'u^2 + \mathcal{B}'v^2 + \mathcal{C}'w^2, \\ T^{(4)} &= \mathcal{D}'u^4 + \mathcal{E}'v^4 + \mathcal{F}'w^4 + \mathcal{G}'u^2v^2 + \mathcal{H}'u^2w^2 + \mathcal{K}'v^2w^2, \end{aligned}\right\} \quad (47)$$

где

$$\left.\begin{aligned} \mathcal{A}' &= \frac{1}{2n}\left[-\frac{1}{2}r + a_0\right], \\ \mathcal{B}' &= \frac{1}{2n}\left[-\frac{1}{2}r + a_1\right], \\ \mathcal{C}' &= \frac{r}{2n}, \\ \mathcal{D}' &= -\frac{r}{16n^3}\left[-\frac{1+b}{4} + 1\right] + \frac{a_0}{8n^3}, \\ \mathcal{E}' &= -\frac{r}{16n^3}\left[-\frac{1+b}{4} + 1\right] + \frac{a_1}{8n^3}, \\ \mathcal{F}' &= \frac{(1+b)\,r}{16n^3}, \\ \mathcal{G}' &= \frac{r}{16n^3}\left[\frac{1+b}{2} - 2\right], \\ \mathcal{H}' &= \frac{r}{8n^3}\left[-\frac{1+b}{2} + 1\right], \\ \mathcal{K}' &= \frac{r}{8n^3}\left[-\frac{1+b}{2} + 1\right]. \end{aligned}\right\} \quad (48)$$

§ 4.2. Идеальное отображение

Рассмотрим распространение света от точечного источника, расположенного в точке P_0, в среде с показателем преломления $n(x, y, z)$ Из P_0 выходит бесконечное число лучей, но, вообще говоря, через любую другую точку среды проходит конечное их число Однако в некоторых частных случаях удается найти такую точку P_1, через которую тоже проходит бесконечное количество лучей Эта точка P_1 называется *стигматическим* (или *резким*) изображением P_0.

Идеальный оптический прибор дает для каждой точки P_0 трехмерной области, называемой *пространством предмета*, стигматическое изображение P_1 Совокупность точек изображений определяет *пространство изображения* Соответствующие точки в этих двух областях называются *сопряженными* В общем случае не все лучи, выходящие из P, достигнут пространства изображения, например, некоторые лучи задержатся диафрагмой прибора Говорят, что лучи, достигающие пространства изображения, *лежат в поле прибора* Если P_0 описывает в пространстве предмета кривую C_0, то P_1 описывает сопряженную кривую C_1 Обе кривые не обязательно геометрически подобны друг другу Если *каждая* кривая C_0 в пространстве предмета геометрически подобна своему изображению, то говорят, что эти области *идеально* отображают друг друга Таким же образом можно определить идеальное отображение и для поверхностей

Оптические приборы, дающие идеальные изображения в указанном смысле, представляют значительный интерес, поэтому мы сформулируем некоторые общие теоремы, относящиеся к идеальному, или по крайней мере резкому, отображению трехмерных предметов В п 4 2 3 будут кратко изложены некоторые результаты исследования резкого отображения двумерных предметов (поверхностей)

4.2.1. Общие теоремы. Оптическая система $\mathcal{Y}$, дающая стигматическое изображение трехмерного предмета, часто называется *абсолютным прибором* Будет показано, что *в абсолютном приборе оптическая длина любой кривой, лежащей в пространстве изображения, равна оптической длине своего изображения* Впервые эта теорема была сформулирована в 1858 г Максвеллом [9, 10] для частного случая, когда и пространство предмета, и пространство изображения однородны Позднее Брунс [11], Клейн [5, 12] (см также [13]) и Либман [14] дали более строгое доказательство *)

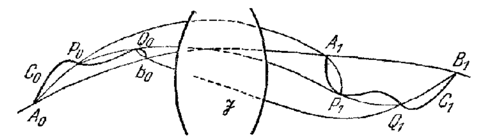


Рис 4 6 Абсолютный оптический прибор.

Как было потом показано Каратеодори [16], эта теорема оказывается справедливой не только для однородной, но и для неоднородной анизотропной среды. При доказательстве ее мы используем метод Каратеодори, однако ограничимся рассмотрением абсолютных приборов с изотропными (но в общем случае неоднородными) пространствами предмета и изображения **).

Пусть A_0B_0 и A_1B_1— отрезки луча в пространстве предмета и в пространстве изображения (рис 4 6), находящегося в поле абсолютного прибора $\mathcal{Y}$ Любой другой луч, линейный элемент которого не отклоняется существенно от элемента $A_0B_0A_1B_1$ ни по абсолютной величине, ни по направлению, также лежит в поле этого прибора

Если каждый элемент кривой (которая, как предполагается, имеет непрерывно поворачивающуюся касательную) совпадает с элементом некоторого луча, полностью лежащего в поле прибора $\mathcal{Y}$, то говорят, что эта кривая расположена в поле $\mathcal{Y}$ *тангенциально* Если заменить ее ломаной с достаточным числом отрезков, то каждый ее отрезок будет совпадать с элементом луча, полностью лежащим в поле прибора

Согласно принципу равного оптического пути (см. п. 3.3.3) все лучи, соединяющие точку A_0 с ее изображением A_1, имеют одинаковую оптическую длину.

*) Обзор этих исследований можно также найти в статье [15]

**) В обычной оптике не используются неоднородные материалы, однако в микроволновой оптике с такими материалами часто имеют дело (см , например, [17]) В электронной оптике «показатель преломления» (см приложение 2) также является функцией координат. Наши рассуждения поэтому представляют не только чисто академический интерес.

Обозначим ее через $V(A_0)$ и покажем, что фактически она не зависит от A_0.

Пусть B_0 и B_1— другая пара сопряженных точек. Тогда (см. рис. 4.6)

$$[A_1B_1]=[A_0B_0]+V(B_0)-V(A_0). \tag{1}$$

Пусть C_0— кривая, соединяющая A_0 и B_0 и лежащая тангенциально в поле прибора, а C_1— ее изображение. Заменим C_0 ломаной $A_0P_0Q_0B_0$ и обозначим изображения точек P_0 и Q_0 через P_1 и Q_1. Применяя (1) к отрезку ломаной A_1P_1, получим

$$[A_1P_1]=[A_0P_0]+V(P_0)-V(A_0),$$

и аналогично для других отрезков

$$[P_1Q_1]=[P_0Q_0]+V(Q_0)-V(P_0) \quad \text{и} \quad [Q_1B_1]=[Q_0B_0]+V(B_0)-V(Q_0).$$

Следовательно,

$$[A_1P_1]+[P_1Q_1]+[Q_1B_1]=[A_0P_0]+[P_0Q_0]+[Q_0B_0]+V(B_0)-V(A_0).$$

Очевидно, что этот результат можно обобщить на случай ломаной с любым числом отрезков N. Переходя к пределу при $N\to\infty$ так, чтобы длина наибольшего отрезка стремилась к нулю, получим соотношение

$$L_1=L_0+V(B_0)-V(A_0), \tag{2}$$

где

$$L_0=\int\limits_{C_0} n_0\,ds_0, \qquad L_1=\int\limits_{C_1} n_1\,ds_1 \tag{3}$$

являются оптическими длинами кривых C_0 и C_1. Покажем теперь, что $V(B_0)=V(A_0)$.

Точки на кривых C_0 и C_1 находятся во взаимно однозначном соответствии, что можно выразить следующими соотношениями:

$$x_1=f(x_0,\ y_0,\ z_0), \qquad y_1=g(x_0,\ y_0,\ z_0), \qquad z_1=h(x_0,\ y_0,\ z_0). \tag{4}$$

Элемент ds_1 кривой C_1 является функцией соответствующего элемента ds_0, т. е.

$$ds_1=\sqrt{\left(\frac{dx_1}{ds_0}\right)^2+\left(\frac{dy_1}{ds_0}\right)^2+\left(\frac{dz_1}{ds_0}\right)^2}\,ds_0. \tag{5}$$

Следовательно,

$$L_1=\int\limits_{C_1} F\left(x_1, y_1, z_1, \frac{dx_1}{ds_0},\ \frac{dy_1}{ds_0},\ \frac{dz_1}{ds_0}\right)ds_0, \tag{6}$$

где

$$F\left(x_1,\ y_1,\ z_1,\ \frac{dx_1}{ds_0},\ \frac{dy_1}{ds_0},\ \frac{dz_1}{ds_0}\right)=n_1(x_1,\ y_1,\ z_1)\sqrt{\left(\frac{dx_1}{ds_0}\right)^2+\left(\frac{dy_1}{ds_0}\right)^2+\left(\frac{dz_1}{ds_0}\right)^2}$$

— однородная функция первой степени по производным dx_1/ds_0, dy_1/ds_0 и dz_1/ds_0; более того, F не меняется при замене dx_1/ds_0, ... на $-dx_1/ds_0$, ... Далее из (4) имеем

$$\frac{dx_1}{ds_0}=\frac{\partial f}{\partial x_0}\frac{dx_0}{ds_0}+\frac{\partial f}{\partial y_0}\frac{dy_0}{ds_0}+\frac{\partial f}{\partial z_0}\frac{dz_0}{ds_0}, \tag{7}$$

аналогичные соотношения получаются для dy_1/ds_0 и dz_1/ds_0. Следовательно, используя (7) и (4), получим для F следующее выражение:

$$F\left(x_1,\ y_1,\ z_1,\ \frac{dx_1}{ds_0},\ \frac{dy_1}{ds_0},\ \frac{dz_1}{ds_0}\right)=\Phi\left(x_0,\ y_0,\ z_0,\ \frac{dx_0}{ds_0},\ \frac{dy_0}{ds_0},\ \frac{dz_0}{ds_0}\right), \tag{8}$$

где Φ — также однородная функция первой степени по dx_0/ds_0, ...; более того, Φ не меняется при замене dx_0/ds_0, ... на $-dx_0/ds_0$, ..., т. е.

$$\Phi\left(x_0,\ y_0,\ z_0,\ -\frac{dx_0}{ds_0},\ -\frac{dy_0}{ds_0},\ -\frac{dz_0}{ds_0}\right)=\Phi\left(x_0,\ y_0,\ z_0,\ \frac{dx_0}{ds_0},\ \frac{dy_0}{ds_0},\ \frac{dz_0}{ds_0}\right). \tag{9}$$

Из (2), (6) и (8) имеем

$$\int_{C_0} (n_0 - \Phi)\, ds_0 = V(A_0) - V(B_0); \tag{10}$$

это свидетельствует о том, что величина криволинейного интеграла в (10) зависит только от положения крайних точек A_0, B_0 и не зависит от выбора кривой C_0. Кривая C_0, однако, не является совершенно произвольной, поскольку она должна лежать тангенциально в поле прибора. Тем не менее мы вправе заключить, что выражение $(n_0 - \Phi)\, ds_0$ есть полный дифференциал от некоторой функции Ψ, т. е.

$$n_0 - \Phi = \frac{\partial \Psi}{\partial x_0}\frac{dx_0}{ds_0} + \frac{\partial \Psi}{\partial y_0}\frac{dy_0}{ds_0} + \frac{\partial \Psi}{\partial z_0}\frac{dz_0}{ds_0}.$$

Если заменить здесь производные $dx_0/ds_0, \ldots$ на $-dx_0/ds_0, \ldots$, то правая часть изменит знак, а левая, согласно (9), не изменит, что возможно только в том случае, если и левая и правая части этого выражения равны нулю ($\Psi = \text{const}$), следовательно,

$$\Phi = n_0. \tag{11}$$

Тогда из соотношения (10) вытекает, что $V(A_0) = V(B_0)$, и выражение (2) принимает вид $L_1 = L_0$. Следовательно, для любой кривой, имеющей изображение, независимо от того, лежит она в поле прибора тангенциально или нет, справедливо соотношение

$$\int_{C_0} n_0\, ds_0 = \int_{C_1} n_1\, ds_1. \tag{12}$$

Это и есть *теорема Максвелла для абсолютного прибора.*

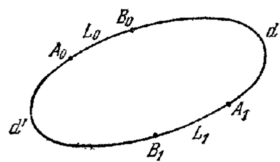


Рис. 4.7. К доказательству Ленца теоремы Максвелла для абсолютного прибора.

Ленц [18] дал следующее менее общее, но чрезвычайно простое и изящное доказательство теоремы Максвелла

Пусть все лучи от каждой точки предмета достигают изображения, и пусть (A_0, A_1) и (B_0, B_1) — две пары сопряженных точек. Согласно предположению луч A_0B_0 должен пройти через A_1 и B_1. Луч B_0A_0 тоже должен пройти через эти точки. Следовательно, каждый луч должен представлять замкнутую кривую. тогда в соответствии с принципом равного оптического пути получим (рис 4 7)

$$[A_0A_1]_{\text{по час. стрелке}} = [A_0A_1]_{\text{против час. стрелки}}$$

и

$$[B_0B_1]_{\text{по час. стрелке}} = [B_0B_1]_{\text{против час. стрелки}}$$

Если ввести обозначения

$$[A_0B_0] = L_0, \qquad [A_1B_1] = L_1, \qquad [B_0A_1] = d, \qquad [B_1A_0] = d',$$

то написанные выше соотношения принимают вид

$$L_0 + d = d' + L_1, \qquad d + L_1 = L_0 + d',$$

откуда следует, что

$$L_0 = L_1.$$

Приведенное здесь доказательство справедливо лишь в том случае, когда кривая является частью луча. Его можно обобщить на случай произвольной кривой таким же способом, как это было сделано нами ранее, т е заменить кривую многоугольником, стороны которого состоят из отрезков лучей, а затем устремить число этих сторон к бесконечности.

Из теоремы Максвелла сразу же вытекает несколько интересных следствий. Рассмотрим небольшой треугольник со сторонами $ds_0^{(1)}$, $ds_0^{(2)}$ и $ds_0^{(3)}$, и пусть $ds_1^{(1)}$, $ds_1^{(2)}$ и $ds_1^{(3)}$ — стороны треугольника, являющегося изображением исходного в абсолютном приборе. Далее, пусть n_0 и n_1 — показатели преломления областей, в которых расположены выбранные треугольники. Согласно теореме Максвелла

$$n_0\, ds_0^{(1)} = n_1\, ds_1^{(1)}, \qquad n_0\, ds_0^{(2)} = n_1\, ds_1^{(2)}, \qquad n_0\, ds_0^{(3)} = n_1\, ds_1^{(3)}, \tag{13}$$

и, следовательно, эти треугольники подобны, а отношения соответствующих сторон обратно пропорциональны величинам показателей преломления. Поэтому углы между любыми двумя кривыми в предмете не изменяются в его изображении, т. е. такое отображение должно быть *конформным*. Из общей теоремы Лиувилля (см., например, [19]) следует, что конформным отображением трехмерной области в трехмерную область может быть только проективное преобразование (коллинеация), инверсия *) или их комбинация. Таким образом, мы доказали следующую теорему, впервые сформулированную Каратеодори. *Отображение, создаваемое абсолютным прибором, является либо проективным преобразованием, либо инверсией, либо их комбинацией.*

Рассмотрим теперь вкратце случай, когда отображение не только стигматично, но и идеально, т. е. любой объект преобразуется в объект, геометрически подобный исходному. Ясно, что такое отображение должно быть проективным преобразованием, поскольку оно преобразует линии в линии [20]. Тогда из (13) следует, что *увеличение ds_1/ds_0 для любых двух сопряженных линейных элементов равно отношению показателей преломления n_0/n_1*. В частности, если $n_0=n_1=$ const, то $ds_1/ds_0=1$, таким образом, *идеальное отображение двух однородных областей с одинаковыми показателями преломления друг в друга всегда тривиально в том смысле, что изображение полностью конгруэнтно предмету.* Единственным известным прибором, обеспечивающим такое отображение, является плоское зеркало (или комбинация плоских зеркал).

Из этих общих рассуждений следует, что для получения нетривиального отображения однородных областей с одинаковыми показателями преломления друг в друга нельзя требовать строгого стигматизма или полного подобия изображения объекту.

4.2.2. «Рыбий глаз» Максвелла. Простым и интересным примером абсолютного прибора можно считать среду, характеризующуюся показателем преломления

$$n(r)=\frac{1}{1+(r/a)^2}n_0, \tag{14}$$

где r — расстояние до фиксированной точки O, а n_0 и a — постоянные. Такая среда называется «рыбьим глазом», и ее свойства были впервые исследованы Максвеллом **) [10, 21].

В § 3.2 было показано, что в сферически симметричной среде лучи представляют собой плоские кривые и лежат в плоскостях, проходящих через начало координат, а уравнение таких лучей имеет вид (см. (3.2.11))

$$\theta=c\int^{r}\frac{dr}{r\sqrt{n^2(r)\,r^2-c^2}},$$

где c — постоянная. Подставляя сюда выражение для $n(r)$ из (14) и полагая

$$\rho=r/a, \qquad K=c/an_0, \tag{15}$$

получим

$$\theta=\int^{\rho}\frac{K(1+\rho^2)\,d\rho}{\rho\sqrt{\rho^2-K^2(1+\rho^2)^2}}. \tag{16}$$

*) При инверсии каждая точка P_0 преобразуется в такую точку P_1, лежащую на линии, которая соединяет P_0 с фиксированным началом координат O, что произведение $OP_0\cdot OP_1$ остается постоянным.

**) Ленц [18] и Стеттлер [22] дали интересные обобщения работы Максвелла. Статья Стеттлера содержит также обобщение теории так называемой *линзы Лунеберга*, которая благодаря своей всенаправленности нашла широкое применение при конструировании антенн сантиметрового диапазона. Такая линза, впервые исследованная Лунебергом [23], представляет собой неоднородную сферу с показателем преломления $n(r)=\sqrt{2-r^2}$ $(0\leqslant r\leqslant 1)$. Если поместить ее в однородную среду с показателем преломления, равным 1, то каждый падающий пучок параллельных лучей соберется в резкий фокус. См. [24—26].

Легко показать, что

$$\frac{K(1+\rho^2)}{\rho\sqrt{\rho^2-K^2(1+\rho^2)^2}}=\frac{d}{d\rho}\left[\arcsin\left(\frac{K}{\sqrt{1-4K^2}}\,\frac{\rho^2-1}{\rho}\right)\right];$$

тогда (16) можно переписать в виде

$$\sin(\theta-\alpha)=\frac{c}{\sqrt{a^2n_0^2-4c^2}}\,\frac{r^2-a^2}{ar}, \qquad (17)$$

где α — постоянная интегрирования.

Соотношение (17) является уравнением лучей в полярных координатах. Поэтому для однопараметрического семейства лучей, проходящих через заданную точку $P_0(r_0, \theta_0)$, имеем

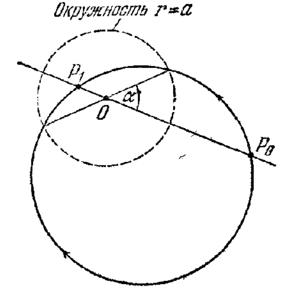


Рис. 4.8. Лучи в «рыбьем глазе» Максвелла.

$$\frac{r^2-a^2}{r\sin(\theta-\alpha)}=\frac{r_0^2-a^2}{r_0\sin(\theta_0-\alpha)}. \qquad (18)$$

Как мы видим, при любом значении параметра α этому уравнению удовлетворяют $r=r_1$, $\theta=\theta_1$, где

$$r_1=a^2/r_0, \quad \theta_1=\pi+\theta_0; \qquad (19)$$

отсюда следует, что *все лучи из произвольной точки P_0 собираются в точке P_1, лежащей на прямой, которая соединяет P_0 и O, P_0 и P_1 расположены по разные стороны от O, и произведение $OP_0\cdot OP_1=a^2$.* Следовательно, «рыбий глаз» служит абсолютным прибором, в котором отображение осуществляется преобразованием *инверсии.*

Отметим, что (17) удовлетворяется при $r=a$, $\theta=\alpha$ и $r=a$, $\theta=\pi+\alpha$, т. е. каждый луч пересекает фиксированную окружность $r=a$ в противоположных точках ее диаметра (рис. 4.8).

Чтобы получить уравнение лучей в декартовых координатах, положим в (17) $x=r\cos\theta$ и $y=r\sin\theta$; тогда

$$y\cos\alpha-x\sin\alpha=\frac{c}{a\sqrt{a^2n_0^2-4c^2}}(x^2+y^2-a^2),$$

или

$$(x+b\sin\alpha)^2+(y-b\cos\alpha)^2=a^2+b^2, \qquad (20)$$

где

$$b=\frac{a}{2c}\sqrt{a^2n_0^2-4c^2}.$$

Из соотношения (20) видно, что каждый луч представляет собой окружность.

4.2.3. Стигматическое отображение поверхностей. До сих пор мы рассматривали только идеальное, или резкое, изображение трехмерных областей. Было показано, что если пространство предмета и пространство изображения однородны и обладают одинаковыми показателями преломления, то идеальное отображение должно быть только тривиальным, т. е. зеркальным отображением предмета. Естественно, возникает вопрос, получается ли нетривиальное отображение, если потребовать, чтобы прибор обеспечивал идеальное (или по крайней мере резкое) отображение лишь некоторых *поверхностей.* Этим вопросом занимался ряд авторов [27—30], которые нашли, что *если пространства предмета и изображения однородны, то вращательно симметричная оптическая система в общем случае*) может обеспечить резкое отображение не более чем*

*) «В общем случае» вставлено для того, чтобы исключить некоторые вырожденные случаи, когда идеально отображается *все пространство предмета* (например, при отражении от плоского зеркала).

двух поверхностей. Доказательство этой теоремы можно найти в работах [27, 29]. Здесь мы подробно рассмотрим только простой случай резкого отображения сферической поверхности, представляющий наибольший практический интерес.

Исследуем преломление лучей на поверхности твердой однородной сферы S, находящейся в однородной среде. Пусть O — точка, где расположен центр сферы, r — ее радиус и n и n' — показатели преломления соответственно материала сферы и окружающей ее среды. Далее пусть AQ — луч, падающий на сферу. С помощью следующих рассуждений легко определить направление преломленного луча QB.

Пусть S_0 и S_1 — сферы, центры которых находятся в O, а радиусы равны

$$r_0 = \frac{n'}{n} r, \qquad r_1 = \frac{n}{n'} r. \tag{21}$$

Если P_0 — точка пересечения AQ с S_0 и P_1 — точка, в которой пересекаются

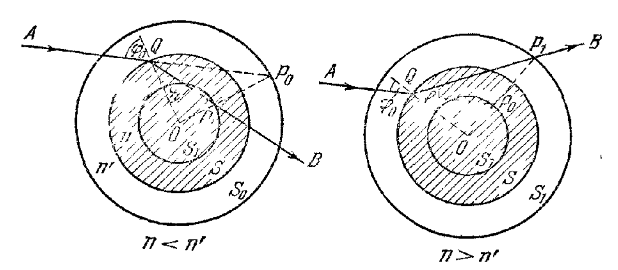


Рис 4 9 Преломление на сферической поверхности. Апланатические точки.

OP_0 и S_1, то QP_1 является преломленным лучом. Последнее следует из построения (рис. 4 9), так как

$$\frac{OQ}{OP_0} = \frac{OP_1}{OQ} = \frac{n}{n'}. \tag{22}$$

Кроме того,

$$\angle QOP_0 = \angle QOP_1. \tag{23}$$

Таким образом, треугольники OQP_0 и OP_1Q подобны, и, следовательно,

$$\frac{\sin\varphi_0}{\sin\varphi_1} = \frac{OP_0}{OQ} = \frac{n'}{n}, \tag{24}$$

где $\varphi_0 = \angle OQP_0$ и $\varphi_1 = \angle OQP_1$ — соответственно углы падения и преломления. Величины углов φ_0 и φ_1 удовлетворяют закону преломления, и поэтому прямая QP_1 представляет собой преломленный луч.

Из построения видно, что все лучи, выходящие из точки P_0 на S_0, дают (виртуальное) стигматическое изображение в точке P_1, расположенной в месте пересечения радиуса OP_0 со сферой S_1. Следовательно, *сфера S_1 представляет собой стигматическое изображение S_0 и наоборот.*

Удобно выразить (24) в несколько иной форме. Если обозначить через θ_0 и θ_1 углы, которые сопряженные лучи образуют с линией P_0P_1, т. е. $\theta_0 = \angle OP_0Q$ и $\theta_1 = \angle OP_1Q$, то, поскольку рассмотренные выше треугольники подобны, $\theta_0 = \varphi_1$ и $\theta_1 = \varphi_0$; следовательно,

$$\frac{\sin\theta_1}{\sin\theta_0} = \frac{n'}{n} = \mathrm{const}. \tag{25}$$

Уравнение (25) является частным случаем так называемого *условия синусов*, значение которого будет объяснено в § 4.5. Согласно терминологии, принятой

в § 4.5, точки P_0 и P_1 называются *апланатическими точками* сферической поверхности S.

Как будет показано в § 6.6, при конструировании некоторых видов объективов микроскопов используется существование апланатических точек преломляющей сферической поверхности.

§ 4.3. Проективное преобразование (коллинеация) при наличии аксиальной симметрии

Выше было показано, что идеальное отображение трехмерных объектов друг в друга можно осуществить лишь с помощью проективного преобразования, поскольку последнее преобразует линии в линии. Свойства проективных преобразований важны даже в тех случаях, когда условия, необходимые для достижения идеального отображения, выполнены не полностью, поскольку, как будет показано позднее, отображение предмета любой оптической системой осуществляется, *по крайней мере в первом приближении*, с помощью преобразования такого рода. Поэтому, прежде чем выводить законы построения изображений в реальных оптических системах, полезно исследовать общие свойства проективных преобразований. Хотя такое предварительное рассмотрение носит чисто геометрический характер, удобно по возможности сохранять оптическую терминологию.

4.3.1. Общие формулы. Пусть (x, y, z) — координаты точки P в пространстве предмета и (x', y', z') — координаты точки P' в пространстве изображения в одной и той же произвольно выбранной декартовой системе координат. Проективное преобразование этих двух областей друг в друга осуществляется с помощью следующих соотношений:

$$x' = F_1/F_0, \quad y' = F_2/F_0, \quad z' = F_3/F_0, \tag{1}$$

где

$$F_i = a_i x + b_i y + c_i z + d_i \qquad (i = 0,\ 1,\ 2,\ 3). \tag{2}$$

Две точки, координаты которых связаны соотношением (1), называются *сопряженной* парой.

Разрешая (1) относительно x, y, z, получим соотношения такого же типа, а именно

$$x = F_1'/F_0', \quad y = F_2'/F_0', \quad z = F_3'/F_0', \tag{3}$$

где

$$F_i' = a_i' x' + b_i' y' + c_i' z' + d_i'.$$

Из (1) следует, что изображение любой точки, лежащей в плоскости $F_0 = 0$, находится на бесконечности. Аналогично из (3) вытекает, что все точки предмета, изображения которых лежат в плоскости $F_0' = 0$, расположены на бесконечности. Плоскость $F_0 = 0$ называется *фокальной плоскостью пространства предмета*, а плоскость $F_0' = 0$ — *фокальной плоскостью пространства изображения* *). Лучи, параллельные в пространстве предмета, преобразуются в лучи, пересекающиеся в некоторой точке, лежащей в фокальной плоскости $F_0' = 0$. Лучи же, выходящие из точки, расположенной в фокальной плоскости $F_0 = 0$, преобразуются в пучок параллельных лучей. В некоторых случаях обе фокальные плоскости находятся на бесконечности. Преобразования такого рода называются *аффинными* или *телескопическими*. При телескопических преобразованиях всегда $F_0 \neq 0$ и $F_0' \neq 0$, так как конечным значением (x, y, z) должны соответствовать конечные значения (x', y', z'). Разумеется, это возможно лишь в том случае, если $a_0 = b_0 = c_0 = 0$ и $a_0' = b_0' = c_0' = 0$.

*) Здесь термины «фокальная плоскость» и «фокальные точки» имеют несколько иной смысл, чем при рассмотрении нормальных конгруэнций (см. п. 3.2.3) и астигматических пучков лучей (см. ниже, § 4.6).

Поскольку большинство оптических систем состоит из поверхностей вращения с общей осью (такие системы обычно называются *центрированными*), особую роль в оптике играет случай *аксиальной симметрии*. Тогда из симметрии системы следует, что изображение любой точки P_0 лежит в плоскости, проходящей через эту точку и ось симметрии; поэтому при изучении свойств соответствующих проективных преобразований можно ограничиться рассмотрением точек, лежащих в такой *меридиональной плоскости*. Пусть эта плоскость совпадает с плоскостью yz, а ось z направлена вдоль оси симметрии. Тогда точка $(0, y, z)$ в пространстве предмета преобразуется в точку $(0, y', z')$ в пространстве изображения, где

$$y' = \frac{b_2 y + c_2 z + d_2}{b_0 y + c_0 z + d_0}, \qquad z' = \frac{b_3 y + c_3 z + d_3}{b_0 y + c_0 z + d_0}. \tag{4}$$

Из симметрии системы вытекает, что z' не меняется при замене y на $-y$. Это возможно только в том случае, если $b_0 = b_3 = 0$. Далее ясно, что при замене y на $-y$ координата y' меняется на $-y'$, откуда следует, что $c_2 = d_2 = 0$. Следовательно, (4) принимает вид

$$y' = \frac{b_2 y}{c_0 z + d_0}, \qquad z' = \frac{c_3 z + d_3}{c_0 z + d_0}. \tag{5}$$

Эти уравнения содержат пять постоянных, однако важны только их отношения. Таким образом, *проективные преобразования при наличии аксиальной симметрии характеризуются четырьмя параметрами.*

Разрешая (5) относительно x и y, получим

$$y = \frac{c_0 d_3 - c_3 d_0}{b_2} \frac{y'}{c_0 z' - c_3}, \qquad z = \frac{-d_0 z' + d_3}{c_0 z' - c_3}. \tag{6}$$

Из соотношений (5) и (6) вытекает, что фокальные плоскости описываются

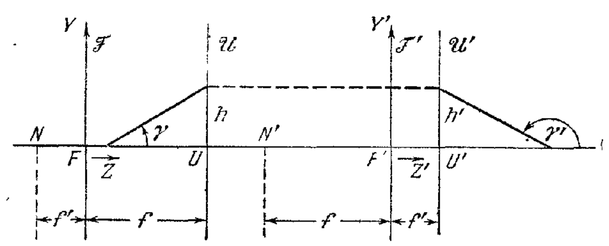


Рис. 4.10. Кардинальные точки и плоскости оптической системы.
F, F' — главные фокусы, U, U' — главные, или единичные, точки; N, N' — узловые точки; $\mathcal{F}$, $\mathcal{F}'$ — фокальные плоскости; $\mathcal{U}$, $\mathcal{U}'$ — главные, или единичные, плоскости.

уравнениями $F_0 \equiv c_0 z + d_0 = 0$, $F_0' \equiv c_0 z' - c_3 = 0$; следовательно, они пересекают ось z под прямым углом в точках, координаты которых равны

$$z = -d_0/c_0, \qquad z' = c_3/c_0. \tag{7}$$

Эти точки называются *главными фокусами* и обозначены на рис. 4.10 через F и F'.

Теперь каждое пространство (пространство предмета и пространство изображения) удобно отнести к своей системе координат, поместив их начала в главные фокусы, т. е. положив

$$\left.\begin{aligned} y &= Y, \qquad c_0 z + d_0 = c_0 Z, \\ y' &= Y', \qquad c_0 z' - c_3 = c_0 Z'. \end{aligned}\right\} \tag{8}$$

Тогда выражение (5) принимает вид

$$Y' = \frac{b_2}{c_0} \frac{Y}{Z}, \qquad Z' = \frac{c_0 d_3 - c_3 d_0}{c_0^2 Z}.$$

Обозначим

$$f=\frac{b_2}{c_0}, \qquad f'=\frac{c_0 d_3-c_3 d_0}{b_2 c_0}. \tag{9}$$

В результате уравнения для преобразований запишутся в простой форме, а именно

$$\frac{Y'}{Y}=\frac{f}{Z}=\frac{Z'}{f'}. \tag{10}$$

Второе из этих уравнений, $ZZ'=ff'$, обычно называют *уравнением Ньютона*. Постоянная f называется *фокусным расстоянием в пространстве предмета*, а f' — *фокусным расстоянием в пространстве изображения*.

Для фиксированных плоскостей предмета и изображения находим из (10)

$$\left(\frac{dY'}{dY}\right)_{Z=\text{const}}=\frac{Y'}{Y}=\frac{f}{Z}=\frac{Z'}{f'}. \tag{11}$$

Эта величина называется *поперечным увеличением*. Далее получим следующее выражение для *продольного увеличения*, не зависящее от Y и Y':

$$\frac{dZ'}{dZ}=-\frac{Z'}{Z}=-\frac{ff'}{Z^2}=-\frac{Z'^2}{ff'}. \tag{12}$$

Из (11) и (12) видно, что величины продольного и поперечного увеличения связаны соотношением

$$\frac{dZ'}{dZ}=-\frac{f'}{f}\left(\frac{dY'}{dY}\right)^2_{Z=\text{const}}. \tag{13}$$

Поскольку величина поперечного увеличения зависит только от Z, но не от Y, фигура, расположенная в плоскости, перпендикулярной к оси, преобразуется в фигуру, геометрически подобную исходной.

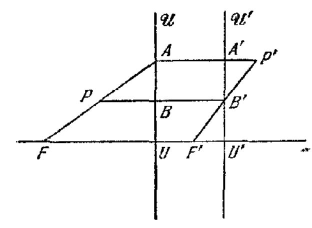


Рис. 4.11. Графическое определение точки изображения

Поперечное увеличение равно единице, если $Z=f$ и $Z'=f'$. Эти плоскости называются *главными* или *единичными плоскостями*, на рис. 4.10 и 4.11 они обозначены буквами $\mathcal{U}$ и $\mathcal{U}'$. Точки их пересечения с осью U и U' называются *главными* или *единичными точками*.

Пусть h — расстояние от оси, на котором луч, выходящий из точки $(0, 0, Z)$, пересекает главную плоскость в пространстве предмета. Сопряженный луч пересекает другую главную плоскость на таком же расстоянии $h'=h$ от оси: углы γ и γ', которые составляют лучи с осью Z, задаются соотношениями (см. рис. 4.10)

$$\operatorname{tg}\gamma=\frac{h}{f-Z}, \qquad \operatorname{tg}\gamma'=\frac{h}{f'-Z'}.$$

Отношение

$$\frac{\operatorname{tg}\gamma'}{\operatorname{tg}\gamma}=\frac{f-Z}{f'-Z'}=-\frac{Z}{f'}=-\frac{f}{Z'} \tag{14}$$

называется *угловым увеличением* или *коэффициентом сходимости*. Эта величина не зависит от h и h' и равна единице, если $Z=-f'$ и $Z'=-f$. Точки на оси N и N', определенные последними двумя равенствами, называются *узловыми*. Они являются сопряженными точками и характеризуются тем, что сопряженные лучи, проходящие через них, параллельны друг другу. Расстояние между узловыми точками равно расстоянию между главными точками. При $f=-f'$ узловые точки совпадают с главными. Фокусы, главные и узловые точки называются *кардинальными* точками данного преобразования.

Иногда оказывается удобным отсчитывать расстояния не от фокальных плоскостей, а от главных. В этом случае вместо Z и Z' используют переменные

$$\zeta = Z - f, \qquad \zeta' = Z' - f'. \tag{15}$$

Уравнение Ньютона $ZZ' = ff'$ принимает тогда вид

$$\frac{f}{\zeta} + \frac{f'}{\zeta'} = -1. \tag{16}$$

Используя свойства фокальных точек и главных плоскостей, можно простым геометрическим построением найти точку P', сопряженную данной точке P. Через точку P проводят две прямые, одна из которых параллельна оси, а другая проходит через фокус F (рис. 4.11). Пусть A и B — точки, в которых эти прямые пересекаются с главной плоскостью $\mathcal{U}$. Тогда из свойств главных плоскостей следует, что точки A' и B', сопряженные A и B, являются точками пересечения двух прямых, проведенных через A и B параллельно оси, с плоскостью $\mathcal{U}'$. Далее, поскольку PA проходит через фокус F, прямая $P'A'$ должна быть параллельна оси, а так как PB параллельна оси, $P'B'$ должна проходить через другой фокус F'. Следовательно, изображение P' находится на пересечении линий AA' и $B'F'$.

4.3.2. Телескопическое отображение. Рассмотрим теперь частный случай телескопического отображения (аффинной коллинеации). Как уже ранее отмечалось, оно характеризуется тем, что обе фокальные плоскости находятся в бесконечности. В этом случае $c_0 = 0$ в (5) и (6) и соотношение (5) принимает вид

$$y' = \frac{b_2 y}{d_0}, \qquad z' = \frac{c_3 z + d_3}{d_0}. \tag{17}$$

Введем опять две самостоятельные координатные системы в пространстве предмета и в пространстве изображения. Начала этих систем поместим в любые две сопряженные точки оси. Формулы преобразования в новых координатах примут простой вид

$$Y' = \alpha Y, \qquad Z' = \beta Z, \tag{18}$$

где $\alpha = b_2/d_0$ и $\beta = c_3/d_0$.

Поперечное и продольное увеличения в данном случае постоянны. Угловое увеличение тоже постоянно: для доказательства последнего утверждения рассмотрим два сопряженных луча, а начала координатных систем поместим в точки, где эти лучи пересекают ось. Тогда уравнения лучей запишутся в виде

$$Y = Z \operatorname{tg} \gamma, \qquad Y' = Z' \operatorname{tg} \gamma'. \tag{19}$$

Следовательно,

$$\frac{\operatorname{tg} \gamma'}{\operatorname{tg} \gamma} = \frac{Y'}{Z'} \frac{Z}{Y} = \frac{\alpha}{\beta} = \frac{b_2}{c_3}. \tag{20}$$

Хотя в случае телескопического отображения как f, так и f' бесконечны, их отношение f'/f должно оставаться конечным. Так, согласно (9) при $c_0 \to 0$ имеем

$$\frac{f'}{f} \to -\frac{c_3 d_0}{b_2^2} = -\frac{\beta}{\alpha^2}. \tag{21}$$

4.3.3. Классификация проективных преобразований. Проективные преобразования можно классифицировать по знаку фокусных расстояний. Если фокусные расстояния имеют разные знаки, то $ff' < 0$, и в соответствии с (12) имеем $dZ'/dZ > 0$. Отсюда следует, что если предмет сместить вдоль оси, то его изображение сместится в том же направлении. Позднее будет показано, что подобное отображение осуществляется при использовании либо только преломления, либо четного числа отражений, либо их комбинации. Такое отображение называется *диоптрическим*.

Если фокусные расстояния имеют одинаковые знаки, то $ff' > 0$, $dZ'/dZ < 0$, и смещению предмета вдоль оси соответствует смещение изображения в противо-

положную сторону. Такое отображение осуществляется с помощью либо нечетного числа отражений, либо комбинации нечетного числа отражений с любым числом преломлений. Подобное отображение называется *катоптрическим*.

В зависимости от знака фокусного расстояния f в каждом из рассмотренных случаев различают два типа преобразований. Из (11) видно, что при $Z > 0$ поперечное увеличение положительно при $f > 0$ и отрицательно при $f < 0$. Следовательно, объект, расположенный в пространстве предмета, имеет прямое или перевернутое изображение, в зависимости от того, положительно ли f или отрицательно. В первом случае преобразование называется *сходящимся*, во втором — *расходящимся*. Такая терминология связана с тем, что падающий пучок параллельных лучей после прохождения главной плоскости в пространстве изображения становится в первом случае сходящимся, а во втором — расходящимся. В табл. 4.1 приведена классификация преобразований.

Таблица 4.1

Классификация проективных преобразований (правило знаков Декарта)

	Сходящиеся	Расходящиеся
Диоптрические	$f > 0;\ f' < 0$	$f < 0;\ f' > 0$
Катоптрические	$f > 0;\ f' > 0$	$f < 0;\ f' < 0$

В частном случае телескопических преобразований указанные четыре типа различают по знакам величин α и β. Из (21) следует, что если $\beta > 0$, то получается диоптрическое преобразование, а если $\beta < 0$, то катоптрическое. Далее из (18) следует, что преобразование будет сходящимся или расходящимся в зависимости от того, положительно ли α или отрицательно.

4.3.4. Комбинация проективных преобразований. Рассмотрим теперь два последовательных проективных преобразования, вращательно симметричных относительно одной и той же оси.

Индексом 0 будем отмечать первое преобразование, а индексом 1 — второе. Тогда уравнения, определяющие эти два преобразования, запишутся в виде

$$\frac{Y_0'}{Y_0}=\frac{f_0}{Z_0}=\frac{Z_0'}{f_0'}, \qquad \frac{Y_1'}{Y_1}=\frac{f_1}{Z_1}=\frac{Z_1'}{f_1'}. \tag{22}$$

Пусть c — расстояние между фокусами F_0' и F_1. Поскольку пространство изображения первого преобразования совпадает с пространством предмета второго, то

$$Z_1=Z_0'-c, \qquad Y_1=Y_0'. \tag{23}$$

Исключая из (22) с помощью (23) координаты промежуточной области, получим

$$\left.\begin{aligned} Y_1'&=\frac{Z_1'Y_1}{f_1'}=\frac{Z_1'Y_0'}{f_1'}=\frac{Z_1'f_0Y_0}{f_1'Z_0}=\frac{f_0f_1Y_0}{f_0f_0'-cZ_0}, \\ Z_1'&=\frac{f_1f_1'}{Z_1}=\frac{f_1f_1'}{Z_0'-c}=\frac{f_1f_1'}{\frac{f_0f_0'}{Z_0}-c}=\frac{f_1f_1'Z_0}{f_0f_0'-cZ_0}. \end{aligned}\right\} \tag{24}$$

Пусть

$$Y=Y_0, \qquad Z=Z_0-\frac{f_0f_0'}{c}, \qquad Y'=Y_1', \qquad Z'=Z_1'+\frac{f_1f_1'}{c}. \tag{25}$$

Уравнения (25) описывают изменение координат при смещении начал координатных систем соответственно на величины $f_0'f_0/c$ и $-f_1f_1'/c$ вдоль оси Z. В этих

координатах уравнения комбинированного преобразования принимают вид

$$\frac{Y'}{Y}=\frac{f}{Z}=\frac{Z'}{f'}, \tag{26}$$

где

$$f=-\frac{f_0 f_1}{c}, \quad f'=\frac{f'_0 f'_1}{c}. \tag{27}$$

Из (25) следует, что расстояния между началами новых и старых систем координат, т. е. расстояния $\delta=F_0F$ и $\delta'=F'_1F'$ между фокусами комбинированного преобразования и фокусами отдельных преобразований, равны соответственно

$$\delta=\frac{f_0 f'_0}{c}, \quad \delta'=-\frac{f_1 f'_1}{c}. \tag{28}$$

Если $c=0$, то $f=f'=\infty$, и эквивалентная коллинеация становится телескопической. Уравнения (24) при этом принимают вид

$$Y'_1=\frac{f_1}{f'_0}Y_0,$$
$$Z'_1=\frac{f_1 f'_1}{f_0 f'_0}Z_0; \tag{29}$$

постоянные α и β из (18) для эквивалентного преобразования выражаются следующим образом:

$$\alpha=\frac{f_1}{f'_0}, \quad \beta=\frac{f_1 f'_1}{f_0 f'_0}. \tag{30}$$

Угловое увеличение равно теперь

$$\frac{\operatorname{tg}\gamma'}{\operatorname{tg}\gamma}=\frac{\alpha}{\beta}=\frac{f_0}{f'_1}. \tag{31}$$

Если одно или оба преобразования телескопические, то приведенные рассуждения необходимо несколько видоизменить.

§ 4.4. Параксиальная оптика

Перейдем теперь к изучению простейших свойств линз, зеркал и их комбинаций. В приводимой ниже теории рассматриваются лишь точки и лучи, лежащие в непосредственной близости от оси; члены, содержащие квадраты и более высокие степени расстояний от оси или углов между лучами и осью, отбрасываются. Эта теория называется параксиальной оптикой или оптикой Гаусса *).

4.4.1. Преломляющая поверхность вращения. Рассмотрим пучок параллельных лучей, падающий на преломляющую поверхность вращения, которая разделяет две однородные среды с показателями преломления n_0 и n_1. Положение точек и лучей в обеих средах будет описываться относительно одной декартовой системы координат, начало которой находится в полюсе O поверхности, а ось z направлена вдоль оси симметрии.

Пусть $P_0(x_0, y_0, z_0)$ и $P_1(x_1, y_1, z_1)$ — точки соответственно на падающем и преломленном лучах. Если пренебречь членами второго и более высоких порядков, то из выражений (4.1.29), (4.1.40) и (4.1.44) следует, что координаты этих точек и лучевые компоненты обоих лучей связаны между собой

*) Здесь, как и раньше, будет использовано обычное правило знаков аналитической геометрии (правило знаков Декарта). Другие правила знаков, применяемые на практике, подробно обсуждаются в докладе, посвященном изучению геометрической оптики и опубликованном Лондонским физическим обществом в 1934 г.

соотношениями

$$\left.\begin{aligned} x_0-\frac{p_0}{n_0}z_0 &= \frac{\partial T^{(2)}}{\partial p_0} = 2\mathcal{A}p_0+\mathcal{C}p_1, \\ x_1-\frac{p_1}{n_1}z_1 &= -\frac{\partial T^{(2)}}{\partial p_1} = -2\mathcal{B}p_1-\mathcal{C}p_0, \end{aligned}\right\} \tag{1а}$$

$$\left.\begin{aligned} y_0-\frac{q_0}{n_0}z_0 &= \frac{\partial T^{(2)}}{\partial q_0} = 2\mathcal{A}q_0+\mathcal{C}q_1, \\ y_1-\frac{q_1}{n_1}z_1 &= -\frac{\partial T^{(2)}}{\partial q_1} = -2\mathcal{B}q_1-\mathcal{C}q_0, \end{aligned}\right\} \tag{1б}$$

где, согласно (4.1.45),

$$\mathcal{A}=\mathcal{B}=\frac{1}{2}\frac{r}{n_1-n_0}, \qquad \mathcal{C}=-\frac{r}{n_1-n_0}, \tag{2}$$

а r — «параксиальный» радиус кривизны поверхности.

Выясним теперь, при каких условиях все лучи, исходящие из точки P_0 (можно предположить, что они лежат в плоскости $x=0$), пройдут после преломления через P_1. В этом случае координаты точки P_1 будут зависеть только от координат P_0 и не будут зависеть от лучевых компонент; поэтому при исключении q_1 из (1б) величина q_0 тоже должна исчезнуть.

Из первого уравнения (1б) следует, что

$$q_1=\frac{1}{\mathcal{C}}\left\{y_0-q_0\left(2\mathcal{A}+\frac{1}{n_0}z_0\right)\right\}. \tag{3}$$

Подставляя это выражение во второе уравнение (1б), получим

$$y_1=-\left(2\mathcal{B}-\frac{1}{n_1}z_1\right)\frac{1}{\mathcal{C}}y_0+\left[\frac{1}{\mathcal{C}}\left(2\mathcal{B}-\frac{z_1}{n_1}\right)\left(2\mathcal{A}+\frac{z_0}{n_0}\right)-\mathcal{C}\right]q_0. \tag{4}$$

Следовательно, P_1 является стигматическим изображением P_0, если

$$\left(2\mathcal{A}+\frac{1}{n_0}z_0\right)\left(2\mathcal{B}-\frac{1}{n_1}z_1\right)=\mathcal{C}^2, \tag{5}$$

или после подстановки сюда (2)

$$\left[\frac{r}{n_1-n_0}+\frac{z_0}{n_0}\right]\left[\frac{r}{n_1-n_0}-\frac{z_1}{n_1}\right]=\frac{r^2}{(n_1-n_0)^2}, \tag{6}$$

что можно записать также в виде

$$n_0\left[\frac{1}{r}-\frac{1}{z_0}\right]=n_1\left[\frac{1}{r}-\frac{1}{z_1}\right]. \tag{7}$$

Из (7) видно, что в рассматриваемом нами приближении каждая точка дает стигматическое изображение; расстояния сопряженных плоскостей до полюса преломляющей поверхности O связаны между собой соотношением (7). Более того, уравнение (4) при условии (5) означает, что данное отображение является *проективным преобразованием.*

Правая и левая части (7) называются *инвариантом Аббе (для преломления)* и играют важную роль в теории оптического отображения. Соотношение (7) можно еще представить в виде

$$\frac{n_1}{z_1}-\frac{n_0}{z_0}=\frac{n_1-n_0}{r}. \tag{8}$$

Величина $(n_1-n_0)/r$ называется *оптической силой* преломляющей поверхности и обозначается буквой $\mathcal{P}$, т. е.

$$\mathcal{P}=\frac{n_1-n_0}{r}. \tag{9}$$

Согласно (4) и (5) поперечное увеличение y_1/y_0 равно единице, если $z_1/n_1=2\mathcal{B}+\mathcal{C}$. Однако из (2) вытекает, что $2\mathcal{B}+\mathcal{C}=0$. Следовательно, координаты z_0 и z_1 главных точек U_0 и U_1 равны нулю, т. е. *главные точки совпадают с полюсом поверхности вращения.* Далее из (8) имеем

$$z_0\to-\frac{n_0r}{n_1-n_0} \text{ при } z_1\to\infty \quad\text{и}\quad z_1\to\frac{n_1r}{n_1-n_0} \text{ при } z_0\to-\infty,$$

следовательно, абсциссы фокусов F_0 и F_1 равны соответственно $-n_0 r/(n_1 - n_0)$ и $n_1 r/(n_1 - n_0)$. Поэтому фокусные расстояния $f_0 = F_0 U_0$ и $f_1 = F_1 U_1$ определяются соотношениями

$$f_0 = \frac{n_0 r}{n_1 - n_0}, \qquad f_1 = -\frac{n_1 r}{n_1 - n_0}, \tag{10}$$

или, используя выражение для $\mathcal{P}$

$$\frac{n_0}{f_0} = -\frac{n_1}{f_1} = \mathcal{P}. \tag{11}$$

Позже будет показано, что соотношение $n_0/f_0 = -n_1/f_1$ справедливо не только для одной преломляющей поверхности, но и для любой центрированной системы, причем величины с индексом 0 относятся к пространству предмета, а величины с индексом 1 — к пространству изображения. Следовательно, можно считать, что соотношение (11) служит общим определением оптической силы центрированной системы. Практической единицей оптической силы является диоптрия, соответствующая силе сферической поверхности с фокусным расстоянием $f_0 = 1$ *м* в вакууме ($n_0 = 1$). Оптическая сила положительна для собирательной системы ($f_0 > 0$) и отрицательна для рассеивающей ($f_0 < 0$)

Поскольку f_0 и f_1 имеют разные знаки, отображение оказывается *диоптрическим* (см. п. 4.3.3). Если поверхность обращена своей выпуклой частью к падающему пучку ($r > 0$) и $n_0 < n_1$, то $f_0 > 0$, $f_1 < 0$, и отображение получается *сходящимся*. Если $r > 0$, $n_0 > n_1$, то оно будет *расходящимся* (рис. 4.12). Если же поверхность обращена своей вогнутой частью к падающему пучку, то типы отображений поменяются местами.

Используя (10), можно записать соотношение (8) в виде

$$\frac{f_0}{z_0} + \frac{f_1}{z_1} = -1, \tag{12}$$

а коэффициенты (2) — в форме

$$\mathcal{A} = \frac{f_0}{2n_0}, \qquad \mathcal{B} = -\frac{f_1}{2n_1}, \qquad \mathcal{C} = -\frac{f_0}{n_0} = \frac{f_1}{n_1}. \tag{13}$$

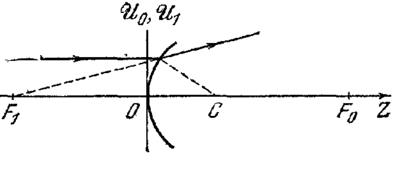


Рис. 4.12. Положение кардинальных точек при преломлении на поверхности вращения.

Введем в каждом из двух пространств свои системы координат, начала которых находятся в фокусах, а оси параллельны осям исходной системы; тогда

$$X_0 = x_0, \qquad X_1 = x_1,$$
$$Y_0 = y_0, \qquad Y_1 = y_1,$$
$$Z_0 = z_0 + f_0, \qquad Z_1 = z_1 + f_1.$$

В этих координатах уравнения (12) и (4) при выполнении условия (5) принимают обычный вид (4.3.10), т. е.

$$\frac{Y_1}{Y_0} = \frac{f_0}{Z_0} = \frac{Z_1}{f_1}. \tag{14}$$

4.4.2. Отражающая поверхность вращения. Легко показать, что поведение световых лучей на отражающей поверхности вращения описывается теми же уравнениями (1а) и (1б), что и выше, если коэффициенты $\mathcal{A}$, $\mathcal{B}$ и $\mathcal{C}$ в них заменить на $\mathcal{A}'$, $\mathcal{B}'$ и $\mathcal{C}'$, введенные в п. 4.1.5. Там же было отмечено, что $\mathcal{A}'$, $\mathcal{B}'$ и $\mathcal{C}'$ получаются из $\mathcal{A}$, $\mathcal{B}$ и $\mathcal{C}$, если положить $n_0 = -n_1 = n$, где n — показа-

тель преломления среды, в которой распространяются лучи. Следовательно, положив $n_0 = -n_1 = n$ в предыдущих формулах, можно сразу же записать соответствующие выражения для отражающей поверхности. В частности, из (7) следует, что

$$\frac{1}{z_0} - \frac{1}{r} = -\frac{1}{z_1} + \frac{1}{r}, \tag{15}$$

где правая и левая части являются *инвариантом Аббе для отражения*. Последнее выражение можно представить в виде

$$\frac{1}{z_0} + \frac{1}{z_1} = \frac{2}{r}. \tag{16}$$

Фокусные расстояния f_0 и f_1 в этом случае равны

$$f_0 = f_1 = -\frac{r}{2}, \tag{17}$$

а оптическая сила —

$$\mathcal{P} = -\frac{2n}{r}. \tag{18}$$

Так как произведение $f_0 f_1$ больше нуля, то получается *катоптрическое* отображение. Если поверхность обращена своей выпуклой частью к падающему пучку ($r > 0$), то $f_0 < 0$ и отображение оказывается расходящимся; если же она обращена к нему своей вогнутой частью ($r < 0$), то $f_0 > 0$ и получается сходящееся отображение (рис. 4.13).

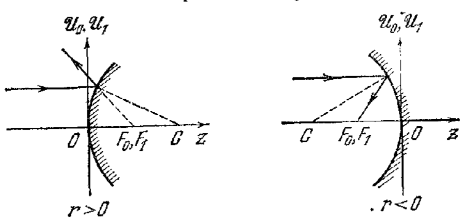


Рис. 4.13. Положение кардинальных точек при отражении от поверхности вращения

4.4.3. Толстая линза. Получим теперь формулы Гаусса для отображения двумя поверхностями, вращательно симметричными относительно одной оси.

Пусть n_0, n_1 и n_2 — показатели преломления трех сред в том порядке, в каком луч пересекает их, и r_1 и r_2 — «параксиальные» радиусы кривизны поверхностей, считающиеся положительными при падении света на выпуклую сторону поверхности. Фокусные расстояния первой поверхности выражаются на основании (10) формулами

$$f_0 = \frac{n_0 r_1}{n_1 - n_0}, \qquad f_0' = -\frac{n_1 r_1}{n_1 - n_0}, \tag{19}$$

а второй поверхности — формулами

$$f_1 = \frac{n_1 r_2}{n_2 - n_1}, \qquad f_1' = -\frac{n_2 r_2}{n_2 - n_1}. \tag{20}$$

Фокусные расстояния комбинации поверхностей, согласно (4.3.27), равны

$$f = -\frac{f_0 f_1}{c}, \qquad f' = \frac{f_0' f_1'}{c}, \tag{21}$$

где c — расстояние между фокусами F_0' и F_1. Пусть t — осевая толщина линзы, т. е. расстояние между полюсами двух поверхностей; тогда (рис. 4.14)

$$c = t + f_0' - f_1. \tag{22}$$

Подставляя сюда выражения для f_0' и f_1, получим

$$c = \frac{D}{(n_1 - n_0)(n_2 - n_1)}, \tag{23}$$

где

$$D = (n_1 - n_0)(n_2 - n_1)\,t - n_1\left[(n_2 - n_1)\,r_1 + (n_1 - n_0)\,r_2\right]. \tag{24}$$

Если мы теперь подставим в (21) выражения для f_0, f_1, f'_0, f'_1 и c, то получим требуемые выражения для фокусных расстояний комбинации двух поверхностей, а именно

$$f=-n_0n_1\frac{r_1r_2}{D},\qquad f'=n_1n_2\frac{r_1r_2}{D}. \tag{25}$$

Поскольку произведение $ff'<0$, отображение будет диоптрическим. Оптическая сила $\mathcal{P}$ линзы равна

$$\mathcal{P}=\frac{n_0}{f}=-\frac{n_2}{f}=-\frac{D}{nr_1r_2}=$$
$$=\mathcal{P}_1+\mathcal{P}_2-\frac{t}{n_1}\mathcal{P}_1\mathcal{P}_2, \tag{26}$$

где $\mathcal{P}_1$ и $\mathcal{P}_2$ — оптические силы двух поверхностей.

Из (4.3.28) следует, что расстояния $\delta=F_0F$ и $\delta'=F'_1F'$ равны

$$\delta=-n_0n_1\frac{n_2-n_1}{n_1-n_0}\frac{r_1^2}{D},$$
$$\delta'=n_1n_2\frac{n_1-n_0}{n_2-n_1}\frac{r_2^2}{D}. \tag{27}$$

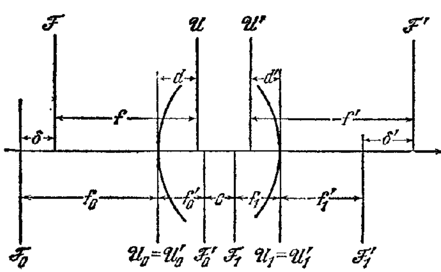


Рис. 4.14. Кардинальные точки комбинированной системы (толстая линза).

Расстояния d и d' от главных плоскостей $\mathcal{U}$ и $\mathcal{U}'$ до полюсов поверхностей выражаются формулами (см. рис. 4.14)

$$\left.\begin{aligned} d&=\delta+f-f_0=-n_0(n_2-n_1)\frac{r_1t}{D},\\ d'&=\delta'+f'-f'_1=n_2(n_1-n_0)\frac{r_2t}{D}.\end{aligned}\right\} \tag{28}$$

Особый интерес представляет случай равенства показателей преломления сред по обе стороны от линзы, т. е. $n_2=n_0$. Положив в предыдущих формулах $n_1/n_0=n_1/n_2=n$, получим

$$\left.\begin{aligned} &f=-f'=-\frac{nr_1r_2}{\Delta},\qquad \delta=\frac{nr_1^2}{\Delta},\qquad \delta'=-\frac{nr_2^2}{\Delta},\\ &d=\frac{(n-1)r_1t}{\Delta},\qquad d'=\frac{(n-1)r_2t}{\Delta},\end{aligned}\right\} \tag{29}$$

где

$$\Delta=(n-1)\left[n(r_1-r_2)-(n-1)t\right]. \tag{30}$$

В системах координат с началами в фокусах F и F' абсциссы главных точек $Z=f$ и $Z'=f'$, а абсциссы узловых точек $Z=-f'$ и $Z=-f$. Поскольку $f=-f'$, главные и узловые точки в этом случае совпадают. Выражение (4.3.16), связывающее расстояния ζ и ζ' между сопряженными и главными плоскостями, принимает вид

$$\frac{1}{\zeta}-\frac{1}{\zeta'}=-\frac{1}{f}. \tag{31}$$

Рассматриваемая линза является *собирательной* ($f>0$) или *рассеивающей* ($f<0$), в соответствии с

$$f=-\frac{nr_1r_2}{\Delta}\gtrless 0, \tag{32}$$

т. е. с

$$\frac{1}{r_2}-\frac{1}{r_1}\lessgtr\frac{n-1}{n}\frac{t}{r_1r_2}. \tag{33}$$

При $f=\infty$ имеет место промежуточный случай *телескопического* отображения.

Тогда $\Delta = 0$, т. е.

$$r_1 - r_2 = \frac{n-1}{n} t. \tag{34}$$

Указанные выше три типа отображений можно проиллюстрировать на примере двояковыпуклой линзы, у которой $r_1 > 0$, а $r_2 < 0$ (рис. 4.15). Если для простоты предположить, что оба радиуса кривизны равны друг другу по абсолютной величине, т. е. $r_1 = -r_2 = r$, то отображение будет сходящимся или расходящимся в соответствии с $t \lessgtr 2nr/(n-1)$, и телескопическим, когда $t = 2nr/(n-1)$.

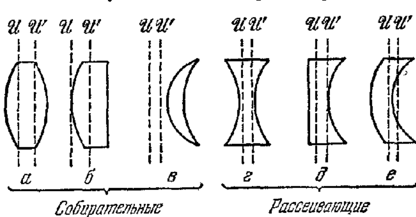


Рис. 4.15. Линзы обычного типа. *а* — двояковыпуклая линза, *б* — плоско выпуклая, *в* — собирательный мениск, *г* — двояковогнутая, *д* — плоско вогнутая, *е* — рассеивающий мениск. $\mathcal{U}$ и $\mathcal{U}'$ — главные плоскости, предполагается, что свет падает слева.

4.4.4. Тонкая линза. Формулы предыдущего параграфа существенно упрощаются, если толщина линзы t настолько мала, что ею можно пренебречь. В этом случае, согласно (28), $d = d' = 0$, т. е. главные плоскости проходят через осевую точку (бесконечно тонкой) линзы. Следовательно, лучи, пересекающие центр линзы, не отклоняются; это означает, что *отображение тонкой линзой является центральной проекцией из центра линзы.*

Положив в (26) $t = 0$, получим

$$\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 = \frac{n_1 - n_0}{r_1} + \frac{n_2 - n_1}{r_2}, \tag{35}$$

т. е. оптическая сила тонкой линзы равна сумме оптических сил двух образующих ее поверхностей.

Если показатели преломления сред, расположенных по обе стороны от линзы, одинаковы ($n_0 = n_2$), то из (25) имеем

$$\frac{1}{f} = -\frac{1}{f'} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right), \tag{36}$$

где, как и раньше, $n = n_1/n_0 = n_1/n_2$. Полагая $n > 1$ (что обычно выполняется), получим, что величина f положительна или отрицательна в зависимости от того, больше или меньше кривизна $1/r_1$ первой поверхности, чем кривизна $1/r_2$ второй (выбор знака связан с кривизной). Отсюда следует, что тонкие линзы, толщина которых уменьшается от центра к краю, являются собирательными, а линзы, толщина которых возрастает к краю,— рассеивающими.

Выпишем выражения, которые понадобятся нам позже для фокусных расстояний f и f' системы из двух центрированных тонких линз, находящихся в воздухе. Согласно (4.3.27) получим, учитывая, что $f_0 = -f'_0$, $f_1 = -f'_1$, соотношение

$$\frac{1}{f} = -\frac{1}{f'} = \frac{c}{f_0 f_1}, \tag{37}$$

где c — расстояние между фокусами F'_0 и F_1 (рис. 4.16). Если l — расстояние между обеими линзами, то

$$l = f_0 + c + f_1. \tag{38}$$

Следовательно,

$$\frac{1}{f} = -\frac{1}{f'} = \frac{1}{f_0} + \frac{1}{f_1} - \frac{l}{f_0 f_1}. \tag{39}$$

Если линзы соприкасаются ($l = 0$), то последнее соотношение можно записать в виде $\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2$, т. е. оптическая сила системы равна в этом случае просто сумме оптических сил двух линз.

4.4.5. Произвольная центрированная система. Было показано, что в приближении параксиальной оптики преломление и отражение лучей на поверхности вращения описывается проективными соотношениями между величинами, относящимися к пространствам предмета и изображения *). Поскольку, согласно § 4.3, последовательное применение нескольких проективных преобразований эквивалентно одному проективному преобразованию, в этом приближении отображение центрированной системой тоже оказывается таким преобразованием. Используя формулы, приведенные в пп. 4.4.1, 4.4.2 и 4.4.4, можно

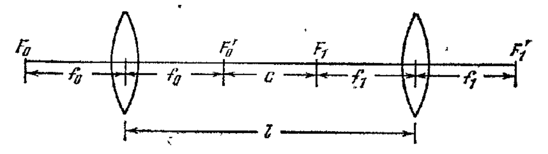


Рис. 4.16. Система из двух центрированных тонких линз.

найти кардинальные точки эквивалентного преобразования. Наше рассмотрение будет в основном посвящено выводу важного инвариантного соотношения, справедливого (в принятом приближении) для любой центрированной системы.

Пусть S_1, S_2, ..., S_m— последовательные поверхности системы, f_0, f_0', f_1, f_1', ..., f_m, f_m'— соответствующие фокусные расстояния и n_0, n_1, ..., n_m— показатели преломления соответствующих сред (рис. 4.17). Далее, пусть P_0,

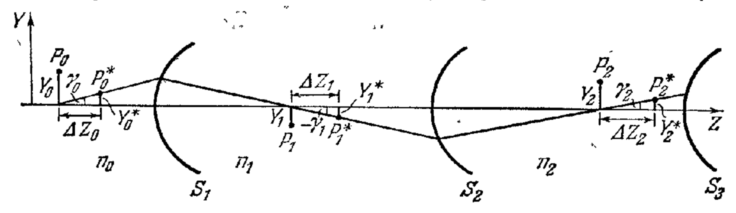


Рис. 4.17. К выводу формулы Смита — Гельмгольца.

P_0^* — две точки в пространстве предмета, расположенные в меридиональной плоскости, а P_1, P_1^*, P_2, P_2^*, ... — их изображения последовательными поверхностями. В системе координат с началом в фокусе первой поверхности координаты точек P_0, P_0^* и P_1, P_1^* связаны соотношениями

$$\frac{Y_1}{Y_0} = \frac{f_0}{Z_0} = \frac{Z_1}{f_1}, \tag{40}$$

$$\frac{Y_1^*}{Y_0^*} = \frac{f_0}{Z_0^*} = \frac{Z_1^*}{f_1}. \tag{41}$$

Следовательно,

$$\frac{Z_1^* - Z_1}{f_1} = \frac{Y_1^*}{Y_0^*} - \frac{Y_1}{Y_0} = \frac{Y_0 Y_1^* - Y_1 Y_0^*}{Y_0 Y_0^*}. \tag{42}$$

*) Как и в случае одной поверхности, пространства предмета и изображения считаются наложенными друг на друга и неограниченно протяженными во всех направлениях. Часть пространства предмета, расположенная перед первой поверхностью (отсчитываемой по ходу световых лучей через систему), называется действительной частью пространства предмета, а часть пространства изображения, расположенная за последней поверхностью, называется действительной частью пространства изображения. Оставшиеся части обоих пространств называются *виртуальными*. Таким же способом можно определить действительную и виртуальную части любой промежуточной области системы.

и

$$\frac{Z_0^*-Z_0}{f_0}=\frac{Y_0^*}{Y_1^*}-\frac{Y_0}{Y_1}=\frac{Y_1Y_0^*-Y_0Y_1^*}{Y_1Y_1^*}. \qquad (43)$$

Пусть

$$Z_0^*-Z_0=\Delta Z_0, \qquad Z_1^*-Z_1=\Delta Z_1. \qquad (44)$$

Тогда из (42) и (43) найдем

$$\frac{f_0Y_0Y_0^*}{\Delta Z_0}=-\frac{f_1Y_1Y_1^*}{\Delta Z_1}. \qquad (45)$$

Согласно (10) $f_0f_1=-n_0/n_1$, так что последнее уравнение принимает вид

$$\frac{n_0Y_0Y_0^*}{\Delta Z_0}=\frac{n_1Y_1Y_1^*}{\Delta Z_1}. \qquad (46)$$

Аналогичным образом получим выражение для преломления на второй поверхности

$$\frac{n_1Y_1Y_1^*}{\Delta Z_1}=\frac{n_2Y_2Y_2^*}{\Delta Z_2}, \qquad (47)$$

т. е. в общем случае

$$\frac{n_{i-1}Y_{i-1}Y_{i-1}^*}{\Delta Z_{i-1}}=\frac{n_iY_iY_i^*}{\Delta Z_i} \quad (1\leqslant i\leqslant m). \qquad (48)$$

Следовательно, $n_iY_iY_i^*/\Delta Z_i$ *есть инвариант последовательных преобразований.* Этот результат играет важную роль в геометрической теории построения изображения. Если положить $Y_i^*/\Delta Z_i=\operatorname{tg}\gamma_i$ (см. рис. 4.17), то (48) примет вид

$$n_{i-1}Y_{i-1}\operatorname{tg}\gamma_{i-1}=n_iY_i\operatorname{tg}\gamma_i.$$

В рассматриваемом нами приближении $\operatorname{tg}\gamma$ и $\operatorname{tg}\gamma'$ можно заменить соответственно на γ и γ', и тогда мы получим *формулу Смита — Гельмгольца* *)

$$n_{i-1}Y_{i-1}\gamma_{i-1}=n_iY_i\gamma_i. \qquad (49)$$

Величина $n_iY_i\gamma_i$ называется *инвариантом Смита — Гельмгольца.*

Из (48) и (49) можно вывести целый ряд важных следствий. Поскольку обычно интересуются только соотношениями между величинами, относящимися к первой и последней средам (пространство предмета и пространство изображения), мы будем в дальнейшем опускать все индексы и обозначать величины, относящиеся к этим двум средам, соответственно нештрихованными и штрихованными символами. Пусть (Y, Z) и $(Y+\delta Y,\ Z+\delta Z)$— две соседние точки в пространстве предмета, а (Y', Z') и $(Y'+\delta Y', Z'+\delta Z')$ — сопряженные им точки. Применяя формулу Смита — Гельмгольца, получим

$$\frac{nY(Y+\delta Y)}{\delta Z}=\frac{n'Y'(Y'+\delta Y')}{\delta Z'}. \qquad (50)$$

В пределе при $\delta Y\to 0$ и $\delta Z\to 0$ находим

$$\frac{dZ'}{dZ}=\frac{n'}{n}\left(\frac{Y'}{Y}\right)^2. \qquad (51)$$

Согласно (4.3.11) $Y'/Y=(dY'/dY)_{Z=\text{const}}$; следовательно, последнее выражение принимает вид

$$\frac{dZ'}{dZ}=\frac{n'}{n}\left(\frac{dY'}{dY}\right)^2_{Z=\text{const}}. \qquad (52)$$

Это соотношение называется *формулой Максвелла для относительного увеличения.* Из нее следует, что *продольное увеличение равно квадрату поперечного увеличения, умноженному на отношение показателей преломления n'/n.* В § 4.3

*) Эту формулу иногда связывают также с именами Лагранжа и Клаузиуса. В более простом виде она была известна еще Гюйгенсу и Котсу (см. [31]).

была выведена аналогичная формула (13), связывающая увеличения и отношение фокусных расстояний. Сравнивая последнюю с (52), получим

$$\frac{f'}{f} = -\frac{n'}{n}, \tag{53}$$

т. е. *отношение фокусных расстояний прибора равно отношению показателей преломления n'/n, взятому со знаком минус.*

Из формулы Смита — Гельмгольца следует также, что

$$\frac{dY'}{dY}\frac{\gamma'}{\gamma} = \frac{n}{n'}, \tag{54}$$

т. е. *произведение поперечного и углового увеличений не зависит от выбора сопряженных плоскостей.*

До сих пор предполагалось, что система состоит только из преломляющих поверхностей. Если же одна из них (скажем, i-я) представляет собой зеркало, то вместо (48) находим

$$\frac{Y_{i-1}Y_{i-1}^*}{\Delta Z_{i-1}} = -\frac{Y_i Y_i^*}{\Delta Z_i},$$

где знак минус появился из-за того, что в случае отражения $f_{i-1}/f_i = 1$, тогда как в случае преломления $f_{i-1}/f_i = -n_{i-1}/n_i$. Поэтому в окончательной формуле нужно заменить n' на $-n'$. Такую же замену следует провести и в более общем случае, когда система содержит нечетное число зеркал; если же система содержит четное число зеркал, то окончательная формула не изменяется.

§ 4.5. Стигматическое отображение пучками с большой угловой апертурой

Законы параксиальной оптики были получены в предположении, что размеры предмета и углы, которые образуют лучи с осью, достаточно малы. Часто приходится иметь дело с оптическими системами и малыми предметами, но большими углами наклона лучей относительно оси. В таких случаях имеются

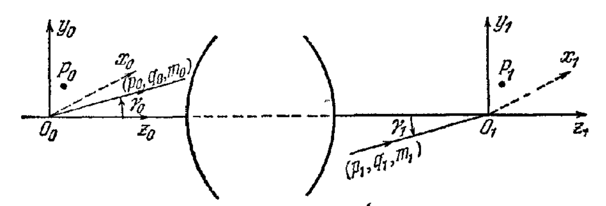


Рис. 4.18. К выводу условия синусов и условия Гершеля.

два простых условия существования стигматического отображения, называемых *условием синусов* *) и *условием Гершеля* [37].

Пусть O_0 — точка предмета, находящаяся на оси, и P_0 — произвольная близкая к ней точка, не обязательно лежащая на оси. Предположим, что система отображает эти точки стигматически и O_1 и P_1 — их изображения.

Пусть (x_0, y_0, z_0) и (x_1, y_1, z_1) — координаты точек P_0 и P_1, причем координаты P_0 берутся относительно прямоугольной системы с центром в O_0, а координаты P_1 — относительно системы с параллельными осями и центром в O_1; оси z в этих системах выбраны так, что они направлены вдоль оптической оси системы (рис. 4.18).

*) Условие синусов было впервые получено Клаузиусом [32] и Гельмгольцем [33] из термодинамических соображений. Однако важность этого условия была замечена лишь после того, как его вторично сформулировал Аббе [34, 35].

В тексте мы следуем, по существу, выводу, предложенному Хокином [36].

Из принципа равного оптического пути вытекает, что длины всех лучей, соединяющих P_0 и P_1, одинаковы. Следовательно, если обозначить через V точечную характеристику среды, то

$$V(x_0, y_0, z_0; x_1, y_1, z_1) - V(0, 0, 0; 0, 0, 0) = F(x_0, y_0, z_0; x_1, y_1, z_1), \qquad (1)$$

где F — некоторая функция, не зависящая от лучевых компонент. Используя основные соотношения (4.1.7), выражающие лучевые компоненты через точечную характеристику, и пренебрегая членами, содержащими вторую и более высокие степени длины, получим из (1)

$$(p_1^{(0)}x_1 + q_1^{(0)}y_1 + m_1^{(0)}z_1) - (p_0^{(0)}x_0 + q_0^{(0)}y_0 + m_0^{(0)}z_0) = F(x_0, y_0, z_0; x_1, y_1, z_1), \qquad (2)$$

где $(p_0^{(0)}, q_0^{(0)}, m_0^{(0)})$ и $(p_1^{(0)}, q_1^{(0)}, m_1^{(0)})$ — лучевые компоненты произвольной пары соответствующих лучей, проходящих через O_0 и O_1. Необходимо отметить, что на величину лучевых компонент не наложено никакого ограничения, хотя вначале и предполагалось, что точки P_0 и P_1 расположены вблизи O_0 и O_1.

Два случая представляют особый интерес, а именно: 1) P_0 и P_1 лежат соответственно в плоскостях $z_0 = 0$ и $z_1 = 0$ и 2) P_0 и P_1 лежат на оси симметрии. Эти два случая будут исследованы отдельно.

4.5.1. Условие синусов. Не уменьшая общности рассуждений, можно и здесь рассматривать только точки, лежащие в меридиональной плоскости ($x_0 = x_1 = 0$). Если P_0 лежит в плоскости $z_0 = 0$, а P_1 — в плоскости $z_1 = 0$, то (2) принимает вид

$$q_1^{(0)}y_1 - q_0^{(0)}y_0 = F(0, y_0, 0; 0, y_1, 0). \qquad (3)$$

Это соотношение справедливо для любой пары сопряженных лучей. Поэтому оно должно быть справедливым и для осевой пары $p_0^{(0)} = q_0^{(0)} = 0$ и $p_1^{(0)} = q_1^{(0)} = 0$. Следовательно,

$$F(0, y_0, 0; 0, y_1, 0) = 0. \qquad (4)$$

Таким образом, (3) запишется следующим образом:

$$q_1^{(0)}y_1 = q_0^{(0)}y_0, \qquad (5)$$

или в явном виде

$$n_1 y_1 \sin\gamma_1 = n_0 y_0 \sin\gamma_0, \qquad (6)$$

где γ_0 и γ_1 — углы, образованные соответствующими лучами, проходящими через O_0 и O_1, с осью z; n_0 и n_1 — показатели преломления среды в пространствах предмета и изображения. Соотношение (6) называется *условием синусов*; это то условие, при выполнении которого небольшая часть плоскости предмета, лежащей вблизи оси, резко отображается расходящимся пучком лучей с любой угловой апертурой. Если угловое расхождение достаточно мало, то $\sin\gamma_0$ и $\sin\gamma_1$ можно заменить соответственно на γ_0 и γ_1, и условие синусов перейдет в формулу Смита — Гельмгольца (4.4.49).

Если предмет находится на бесконечности, то условие синусов имеет другой вид. Предположим сначала, что осевая точка предмета расположена на большом расстоянии от первой поверхности. Если Z_0 — абсцисса этой точки в системе координат с центром в первом фокусе и h_0 — расстояние от оси точки пересечения луча, выходящего из осевой точки, с первой поверхностью, то $Z_0 \sin\gamma_0/h_0 \to -1$ при $Z_0 \to \infty$ и постоянном h_0. Следовательно, если Z_0 достаточно велико, то (6) принимает вид

$$\frac{n_0}{n_1} h_0 = -\frac{y_1}{y_0} Z_0 \sin\gamma_1. \qquad (7)$$

Но согласно (4.3.10) $y_1 Z_0/y_0 = f_0$, а согласно (4.4.53) $n_0/n_1 = -f_0/f_1$, так что в пределе $Z_0 \to \infty$ выражение (6) сводится к (рис. 4.19)

$$\frac{h_0}{\sin\gamma_1} = f_1. \qquad (8)$$

Отсюда следует, что любой луч, падающий параллельно оси, пересечет сопряженный ему луч на сфере радиуса f_1 с центром в фокусе F_1.

Говорят, что две осевые точки образуют *апланатическую* пару, если, во-первых, они являются стигматическими изображениями одна другой и, во-вторых, сопряженные лучи, проходящие через них, удовлетворяют условию синусов. Мы уже встречались с такими точками при изучении преломления лучей на сферической поверхности (см. п. 4.2.3).

В терминах теории аберраций (см. гл. 5) осевой стигматизм означает, что в разложении характеристической функции отсутствуют члены, не зависящие от расстояния предмета до оси, т. е. отсутствует сферическая аберрация всех порядков. Если же выполняется еще и условие синусов, то пропадают члены,

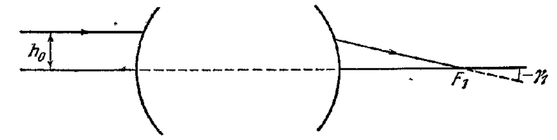


Рис. 4.19. Условие синусов в случае, когда предмет находится в бесконечности.

содержащие расстояние до оси в первой степени; эти члены описывают аберрацию, которая называется *круговой комой*.

Условие синусов играет важную роль при конструировании различных оптических систем, поскольку оно дает информацию о качестве изображения точек, не расположенных на оси, выраженную через свойства аксиальных пучков.

4.5.2. Условие Гершеля. Рассмотрим теперь случай, когда P_0 и P_1 находятся на оси системы ($x_0 = y_0 = 0$, $x_1 = y_1 = 0$). Условие резкого отображения (2) примет тогда вид

$$m_1^{(0)} z_1 - m_0^{(0)} z_0 = F(0,\ 0,\ z_0;\ 0,\ 0,\ z_1), \qquad (9)$$

или через γ_0 и γ_1 —

$$n_1 z_1 \cos\gamma_1 - n_0 z_0 \cos\gamma_0 = F(0,\ 0,\ z_0;\ 0,\ 0,\ z_1). \qquad (10)$$

В частном случае осевого луча получим

$$F(0,\ 0,\ z_0;\ 0,\ 0,\ z_1) = n_1 z_1 - n_0 z_0. \qquad (11)$$

Следовательно, (10) можно представить в виде

$$n_1 z_1 \sin^2\frac{\gamma_1}{2} = n_0 z_0 \sin^2\frac{\gamma_0}{2}. \qquad (12)$$

Это соотношение является одной из форм записи *условия Гершеля*. Поскольку предполагается, что расстояния от начала координат малы, то на основании формулы Максвелла (4.4.52) найдем

$$\frac{z_1}{z_0} = \frac{n_1}{n_0}\left(\frac{y_1}{y_0}\right)^2, \qquad (13)$$

и тогда условие Гершеля можно записать следующим образом:

$$n_1 y_1 \sin\frac{\gamma_1}{2} = n_0 y_0 \sin\frac{\gamma_0}{2}. \qquad (14)$$

Если это условие выполнено, то любой элемент оси, расположенный вблизи O_0, будет резко отображаться пучком лучей независимо от величины его углового расхождения.

Необходимо отметить, что условие синусов и условие Гершеля могут удовлетворяться одновременно лишь в случае $\gamma_1 = \gamma_0$. Тогда $y_1/y_0 = z_1/z_0 = n_0/n_1$, т. е. продольное и поперечное увеличения равны отношению показателей преломления среды в пространстве предмета и в пространстве изображения.

§ 4.6. Астигматические пучки лучей

Прямолинейные лучи, имеющие общую точку, образуют так называемый *гомоцентрический пучок*. Волновые фронты в этом случае имеют вид сферических поверхностей, центры которых находятся в точке пересечения лучей. В предыдущих разделах мы изучали именно такие пучки.

Обычно после преломления или отражения лучей их гомоцентричность нарушается. Поэтому целесообразно исследовать свойства пучков прямолинейных лучей более общего типа.

4.6.1. Фокальные свойства тонкого пучка. Пусть S — один из волновых фронтов пучка прямолинейных лучей, а P — произвольная точка на нем (рис. 4.20). Рассмотрим плоскость, проходящую через луч в точке P, и обозначим кривую пересечения этой плоскости с S через C. Поскольку лучи пучка

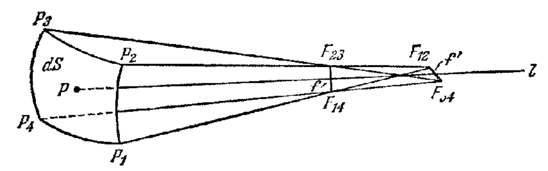


Рис. 4.20 Тонкий пучок лучей.

нормальны к поверхности S, центр круга кривизны в точке P расположен на луче, проходящем через эту точку.

Если теперь поворачивать рассматриваемую плоскость вокруг луча, то кривая C, а следовательно, и радиус кривизны будут изменяться непрерывным образом. При повороте плоскости на 180° величина радиуса кривизны пройдет через свои максимальное и минимальное значения. С помощью простых геометрических рассуждений можно показать *), что две плоскости, в которых лежат максимальный и минимальный радиусы кривизны, взаимно перпендикулярны. Эти плоскости называются *главными плоскостями* **) для точки P, а соответствующие радиусы — *главными радиусами кривизны*. Кривые на поверхности S, касающиеся во всех своих точках главных плоскостей, образуют два взаимно ортогональных семейства кривых, называемых *линиями кривизны*.

В первом приближении две нормали к поверхности в смежных ее точках, вообще говоря, не пересекаются. Однако если эти точки находятся на линии кривизны, то нормали пересекаются и точка их пересечения служит фокусом конгруэнции, образованной нормалями (лучами). Следовательно, в согласии с общими выводами п. 3.2.3 на каждой нормали имеются два фокуса, которые являются главными центрами кривизны. Поэтому каустическая поверхность пучка прямолинейных лучей состоит в общем случае из двух листов и служит *эволютой* волновых фронтов; волновые фронты в свою очередь являются *эвольвентами* каустической поверхности. Если волновые фронты представляют собой поверхности вращения, то один лист каустической поверхности вырождается в отрезок оси вращения, а другой становится поверхностью вращения, меридиональное сечение которой является эволютой меридионального сечения волнового фронта.

Рассмотрим тонкий пучок лучей, пересекающих элемент dS волнового фронта. В качестве границы элемента dS удобно выбрать две пары линий кривизны, которые можно считать отрезками дуг окружностей. Два из них

*) См., например, [38].

**) Термины «главные плоскости» и «фокальные плоскости» имеют здесь другой смысл, чем при изучении проективных преобразований, рассмотренных в § 4.3.

(P_1P_2 и P_4P_3) направлены по вертикали, а два других (P_1P_4 и P_2P_3) — по горизонтали (см. рис. 4.20).

Все лучи, проходящие через дугу P_1P_2, пересекутся (с точностью до членов первого порядка малости) в фокусе F_{12}, а лучи, проходящие через P_3P_4,— в фокусе F_{34}. Линия f', соединяющая F_{12} и F_{34}, называется *фокальной линией* пучка и, как видно из чертежа, лежит в горизонтальной плоскости. Лучам, проходящим через P_1P_4 и P_2P_3, соответствует вертикальная фокальная линия f.

Если провести линии кривизны через любую точку на dS, то соответствующие им фокусы окажутся на двух фокальных линиях, и наоборот. Таким образом, *приближенная модель тонкого пучка лучей получается путем соединения всех пар точек на двух взаимно ортогональных отрезках линий.*

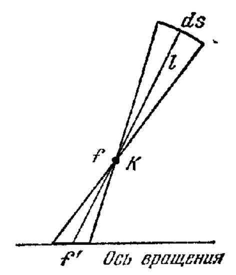


Рис. 4.21 Фокальные линии волнового фронта с цилиндрической симметрией.

Луч l, проходящий через центральную точку P, называется *центральным* (*или главным*) *лучом* пучка, а расстояние между фокальными линиями, измеренное вдоль этого луча,— *астигматической разностью* пучка. Две плоскости, проходящие через f, l и f', l, называются *фокальными плоскостями* пучка; эти плоскости перпендикулярны друг к другу. Однако фокальные линии не обязательно перпендикулярны к центральному лучу (как часто неверно утверждается в литературе). Рассмотрим, например, семейство волновых фронтов, обладающих цилиндрической симметрией относительно общей оси (рис. 4.21). Пусть dS — элемент поверхности (не содержащий осевую точку) одного из волновых фронтов, а ds — участок кривой пересечения dS с плоскостью, проходящей через ось. Тогда ясно, что фокальная линия f в центре кривизны K элемента кривой ds ортогональна этой плоскости. Другая фокальная линия f' совпадает с отрезком оси, ограниченным нормалями к концам ds. В общем случае эта фокальная линия не ортогональна l.

4.6.2. Преломление тонкого пучка. Было показано, что тонкий пучок лучей полностью характеризуется своим центральным лучом и двумя своими

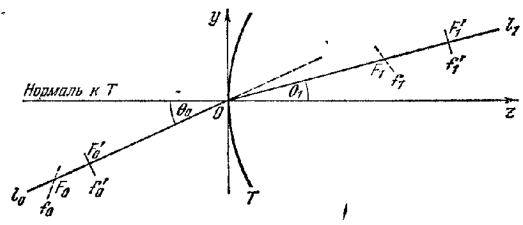


Рис. 4 22. Преломление тонкого астигматического пучка лучей.

фокальными линиями. Предположим, что такой пучок падает на преломляющую поверхность. Определим центральный луч и фокальные линии преломленного пучка. Рассмотрим случай, имеющий большое практическое значение, с именно случай совпадения одной из главных плоскостей падающего пучка а главной плоскостью кривизны преломляющей поверхности в точке O, в которой ее пересекает центральный луч (рис. 4.22).

Выберем декартову систему координат с началом в точке O, осью z, направленной по нормали к поверхности T, и осями x и y, направленными по главным линиям кривизны поверхности T.

Далее, пусть θ_0 и θ_1— углы, образованные центральными лучами l_0 и l_1 двух пучков и осью z, F_0 и F_0'— фокусы падающего пучка, расположенные со-

ответственно при $z=\zeta_0$ и $z=\zeta_1$. Предположим, что фокальная линия в F_0 перпендикулярна к плоскости падения; в этом случае F_0 называется *первичным фокусом*, а соответствующая фокальная линия f_0 — *первичной фокальной линией*. Фокус F_0' называется тогда *вторичным фокусом*, а соответствующая фокальная линия f_0' (лежащая в плоскости падения) — *вторичной фокальной линией*. В случае центрированной системы первичный и вторичный фокусы пучка, центральный луч которого находится в меридиональной плоскости, называют соответственно *тангенциальным и сагиттальным фокусами*.

Для нахождения фокальных линий преломленного пучка необходимо прежде всего выписать выражение для угловой характеристики преломляющей поверхности. Если радиусы кривизны преломляющей поверхности в главных направлениях x и y равны r_x и r_y, то уравнение этой поверхности имеет вид

$$z=\frac{x^2}{2r_x}+\frac{y^2}{2r_y}+\ldots \tag{1}$$

Тогда, согласно (4.1.34), угловая характеристика относительно системы координат с началом в O (при $a_0=a_1=0$) равна

$$T=(p_0-p_1)x+(q_0-q_1)y+(m_0-m_1)z. \tag{2}$$

Из закона преломления, учитывая лишь члены низших порядков, можно получить таким же способом, как и в § 4.1 (где рассматривался случай $r_x=r_y$), соотношения

$$x=-r_x\frac{p_0-p_1}{m_0-m_1},\qquad y=-r_y\frac{q_0-q_1}{m_0-m_1}. \tag{3}$$

Подстановка (3) в (1) дает

$$z=\frac{1}{2\mu^2}\left[r_x(p_0-p_1)^2+r_y(q_0-q_1)^2\right], \tag{4}$$

где

$$\mu=m_1-m_0=n\cos\theta_1-n_0\cos\theta_0. \tag{5}$$

Подставляя теперь (3) и (4) в (2), получим требуемое выражение для угловой характеристики

$$T(p_0,\ q_0;\ p_1,\ q_1)=\frac{1}{2\mu}\left[r_x(p_0-p_1)^2+r_y(q_0-q_1)^2\right]+\ldots \tag{6}$$

Используя это соотношение в (4.1.29), получим уравнения для падающего и преломленного лучей в виде

$$x_0-\frac{p_0}{m_0}z_0=\frac{1}{\mu}r_x(p_0-p_1), \tag{7}$$

$$x_1-\frac{p_1}{m_1}z_1=\frac{1}{\mu}r_y(p_0-p_1); \tag{8}$$

соответствующие уравнения, содержащие координату y, совершенно аналогичны.

Исследуем теперь изменения различных величин при переходе от центрального луча к соседнему. Из (7) и (8) имеем

$$\delta x_0-\frac{p_0}{m_0}\delta z_0=z_0\delta\left(\frac{p_0}{m_0}\right)+\frac{1}{\mu}r_x(\delta p_0-\delta p_1), \tag{9}$$

$$\delta x_1-\frac{p_1}{m_1}\delta z_1=z_1\delta\left(\frac{p_1}{m_1}\right)+\frac{1}{\mu}r_x(\delta p_0-\delta p_1). \tag{10}$$

Лучевые компоненты центрального луча падающего пучка равны

$$p_0=0,\qquad q_0=n_0\sin\theta_0,\qquad m_0=n_0\cos\theta_0, \tag{11}$$

так что

$$\delta\left(\frac{p_0}{m_0}\right)=\frac{1}{m_0^2}(m_0\delta p_0-p_0\delta m_0)=\frac{1}{n_0}\sec\theta_0\delta p_0 \tag{12}$$

и

$$\delta\left(\frac{q_0}{m_0}\right)=\frac{1}{m_0^2}(m_0\delta q_0-q_0\delta m_0)=\frac{1}{n_0}\sec^3\theta_0\delta q_0. \tag{13}$$

Здесь было использовано тождество $m_0\delta m_0+p_0\delta p_0+q_0\delta q_0=0$.

Уравнения (9) и (10) принимают вид

$$\delta x_0=\frac{z_0}{n_0}\sec\theta_0\delta p_0-\frac{1}{\mu}r_x(\delta p_1-\delta p_0), \tag{14}$$

$$\delta x_1=\frac{z_1}{n_1}\sec\theta_1'\delta p_1-\frac{1}{\mu}r_x(\delta p_1-\delta p_0). \tag{15}$$

При выводе (15) использовался тот факт, что $p_1=0$; этот результат следует из закона преломления и предположения, что $p_0=0$. Аналогичным образом получим

$$\delta y_0-(\operatorname{tg}\theta_0)\,\delta z_0=\frac{z_0}{n_0}(\sec^3\theta_0)\,\delta q_0-\frac{1}{\mu}r_y(\delta q_1-\delta q_0), \tag{16}$$

$$\delta y_1-(\operatorname{tg}\theta_1)\delta z_1=\frac{z_1}{n_1}(\sec^3\theta_1)\,\delta q_1-\frac{1}{\mu}r_y(\delta q_1-\delta q_0). \tag{17}$$

Рассмотрим теперь такие лучи пучка, которые проходят через фокус F_0. Тогда $z_0=\zeta_0$, $\delta x_0=\delta y_0=\delta z_0=0$. Поскольку все эти лучи пересекают также фокальную линию f_0', то $\delta p_0=0$. Учитывая это, находим из (14) и (16)

$$\delta p_1=0 \tag{14a}$$

и

$$\frac{\zeta_0}{n_0}\sec^3\theta_0\,\delta q_0-\frac{1}{\mu}r_y(\delta q_1-\delta q_0)=0. \tag{16a}$$

Уравнение (14а) показывает, что соответствующие преломленные лучи лежат в плоскости yz. Так как все лучи из F_0 проходят через фокус $F_1(z_1=\zeta_1)$, то уравнение (17) должно выполняться при $z_1=\zeta_1$, $\delta x_1=\delta y_1=\delta z_1=0$ для любого значения δq_0, т. е.

$$\frac{\zeta_1}{n_1}\sec^3\theta_1\,\delta q_1-\frac{1}{\mu}r_y(\delta q_1-\delta q_0)=0. \tag{17a}$$

Уравнения (16а) и (17а) при произвольном значении δq_0 могут одновременно удовлетворяться только в том случае, если

$$\frac{n_0\cos^3\theta_0}{\zeta_0}-\frac{n_1\cos^3\theta_1}{\zeta_1}=\frac{n_0\cos\theta_0-n_1\cos\theta_1}{r_y}. \tag{18}$$

Это соотношение определяет положение фокуса F_1 преломленных лучей. Из (14а) следует, что фокальная линия, проходящая через F_1, перпендикулярна к плоскости yz и, следовательно, F_1 является *первичным фокусом*.

Для нахождения положения другого фокуса рассмотрим лучи, выходящие из F_0'. В этом случае $z_0=\zeta_0'$, $\delta x_0=\delta y_0=\delta z_0=0$. Поскольку все эти лучи пересекают фокальную линию f_0, имеем $\delta q_0=\delta m_0=0$. Тогда из уравнений (14) и (16) находим

$$\frac{\zeta_0'}{n_0}\sec\theta_0\,\delta p_0-\frac{1}{\mu}r_x(\delta p_1-\delta p_0)=0 \tag{14б}$$

и

$$\delta q_1=0. \tag{16б}$$

Из (16б) следует, что преломленные лучи лежат теперь в плоскости xz. Все эти лучи проходят через второй фокус $F_1'(z_1=\zeta_1')$, так что (15) должно удовлетворяться при $z_1=\zeta_1'$, $\delta x_1=\delta y_1=\delta z_1=0$ для любого значения величины δp_0. Следовательно,

$$\frac{\zeta_1'}{n_1}\sec\theta_1\delta p_1-\frac{1}{\mu}r_x(\delta p_1-\delta p_0)=0. \tag{15б}$$

Поскольку как (15б), так и (14б) должны удовлетворяться при произвольном значении δp_0, то

$$\frac{n_0 \cos\theta_0}{\zeta_0'} - \frac{n_1 \cos\theta_1}{\zeta_1'} = \frac{n_0 \cos\theta_0 - n_1 \cos\theta_1}{r_x}. \tag{19}$$

Полученное соотношение определяет положение *вторичного фокуса* F_1'.

Часто оказывается удобным определять положение фокусов, задавая их расстояния до точки O, а не координаты z. Обозначая $OF_0 = d_0^{(t)}$, $OF_0' = d_0^{(s)}$, $OF_1 = d_1^{(t)}$, $OF_1' = d_1^{(s)}$ (на рис. 4.22 $d_0^{(t)} < 0$, $d_0^{(s)} < 0$, $d_1^{(t)} > 0$, $d_1^{(s)} > 0$), находим

$$\zeta_0 = d_0^{(t)} \cos\theta_0, \quad \zeta_1 = d_1^{(t)} \cos\theta_1, \quad \zeta_0' = d_0^{(s)} \cos\theta_0, \quad \zeta_1' = d_1^{(s)} \cos\theta_1, \tag{20}$$

а уравнения (18) и (19) принимают вид

$$\frac{n_0 \cos^2\theta_0}{d_0^{(t)}} - \frac{n_1 \cos^2\theta_1}{d_1^{(t)}} = \frac{n_0 \cos\theta_0 - n_1 \cos\theta_1}{r_y} \tag{21}$$

и

$$\frac{n_0}{d_0^{(s)}} - \frac{n_1}{d_1^{(s)}} = \frac{n_0 \cos\theta_0 - n_1 \cos\theta_1}{r_x}. \tag{22}$$

Соответствующие соотношения для случая отражения можно получить, если в приведенных выше формулах положить $n_1 = -n_0$.

§ 4.7. Хроматическая аберрация. Дисперсия призмы

В гл. 2 было показано, что показатель преломления не постоянен для данной среды, а зависит от цвета, т. е. от длины волны света. Рассмотрим сейчас некоторые простые следствия из этого, касающиеся работы линз и призм.

4.7.1. Хроматическая аберрация. Если пучок немонохроматического света падает на преломляющую поверхность, то он расщепляется на несколько лучей, каждый из которых имеет определенную длину волны. Поэтому, пересекая оптическую систему, лучи света с различными длинами волн будут

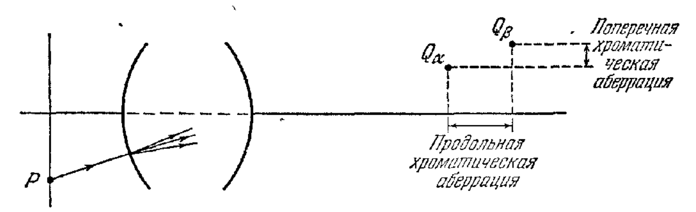


Рис. 4.23. Продольная и поперечная хроматические аберрации.

распространяться после первого преломления не вполне одинаковыми путями. В результате изображение окажется нерезким, и в этом случае говорят, что система обладает *хроматической аберрацией*.

Мы снова ограничимся рассмотрением точек и лучей, расположенных вблизи оси, т. е. предположим, что для каждой длины волны отображение подчиняется законам параксиальной оптики. В этом случае говорят о хроматической аберрации первого порядка, или о первичной аберрации. Пусть Q_α и Q_β — отображения точки P в различных длинах волн (рис. 4.23); тогда проекции $Q_\alpha Q_\beta$ на направления, параллельное и перпендикулярное оси, определяют соответственно *продольную* и *поперечную* хроматические аберрации.

Рассмотрим изменение δf фокусного расстояния тонкой линзы в зависимости от изменения показателя преломления δn. Согласно (4.4.36) величина

$(n-1)f$ для такой линзы не зависит от длины волны. Следовательно,

$$\frac{\delta f}{f}+\frac{\delta n}{n-1}=0. \tag{1}$$

Величина

$$\Delta=\frac{n_F-n_C}{n_D-1}, \tag{2}$$

где n_F, n_D и n_C — показатели преломления, соответствующие линиям Фраунгофера F, D и C (λ4861 Å, 5893 Å и 6563 Å), служит грубой мерой дисперсии стекла и называется *относительной дисперсией*. Из (1) видно, что эта величина приблизительно равна расстоянию между красным и синим изображениями, деленному на фокусное расстояние линзы. На рис. 4.24 показано изменение величин показателей преломления с изменением длины волны для стекла нескольких сортов, обычно используемых в оптических системах. Соответствующие значения Δ лежат в пределах от 1/60 до 1/30.

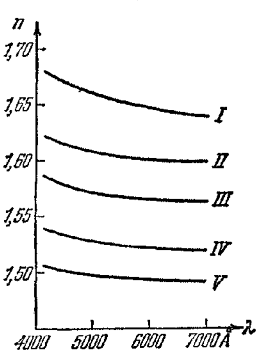


Рис. 4.24. Типичные дисперсионные кривые для стекла различных сортов.

I — тяжелый флинт; *II* — тяжелый бариевый крон; *III* — легкий флинт, *IV* — тяжелый крон; *V* — боросиликатный крон.

Для получения изображения хорошего качества необходимо, чтобы как монохроматические, так и хроматические аберрации были малы. Обычно выбирают некоторое компромиссное решение, поскольку в общем случае невозможно устранить одновременно аберрации всех типов. Часто оказывается достаточным избавиться от хроматической аберрации для двух выбранных длин волн. Выбор этих длин волн зависит, естественно, от назначения той или иной оптической системы; например, фотообъективы, в отличие от приборов, служащих для визуальных наблюдений, обычно «ахроматизируют» для цветов, близких к синему концу спектра, так как обычная фотографическая пластинка более чувствительна к синей области спектра, чем человеческий глаз. Конечно, ахроматизация для двух длин волн не устраняет полностью цветовую ошибку. Остающаяся хроматическая аберрация называется *вторичным спектром*.

Рассмотрим теперь условия, при которых две тонкие линзы образуют комбинацию, свободную от хроматизма фокусного расстояния. Согласно (4.4.39) величина, обратная фокусному расстоянию комбинации двух тонких линз, расположенных на расстоянии l друг от друга, равна

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}-\frac{l}{f_1f_2}. \tag{3}$$

Как мы видим, $\delta f=0$, когда

$$\frac{\delta f_1}{f_1^2}+\frac{\delta f_2}{f_2^2}-\frac{l}{f_1f_2}\left[\frac{\delta f_1}{f_1}+\frac{\delta f_2}{f_2}\right]=0. \tag{4}$$

Если ахроматизация производится для линий C и F, то, используя (1) и (2), получим

$$l=\frac{\Delta_1 f_2+\Delta_2 f_1}{\Delta_1+\Delta_2}, \tag{5}$$

где Δ_1 и Δ_2 — относительные дисперсии обеих линз.

Один из методов уменьшения хроматической аберрации состоит в использовании двух соприкасающихся тонких линз (рис. 4.25), одна из которых сделана из крона, а вторая — из флинта. В этом случае, поскольку $l=0$, получим из (5)

$$\frac{\Delta_1}{f_1}+\frac{\Delta_2}{f_2}=0, \tag{6}$$

или, используя (3),

$$\frac{1}{f_1}=\frac{1}{f}\frac{\Delta_2}{\Delta_2-\Delta_1}, \qquad \frac{1}{f_2}=-\frac{1}{f}\frac{\Delta_1}{\Delta_2-\Delta_1}. \tag{7}$$

Соотношения (7) для данных сортов стекла и заданного фокусного расстояния f однозначно определяют f_1 и f_2. Но f_1 и f_2 зависят от трех радиусов кривизны, следовательно, величину одного из них можно выбрать произвольно. Эта дополнительная степень свободы позволяет иногда уменьшить до минимума сферическую аберрацию.

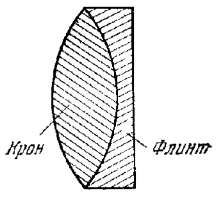


Рис 4 25 Ахроматический дублет

Другой способ создания ахроматической системы состоит в использовании двух тонких линз, изготовленных из одинакового стекла ($\Delta_1=\Delta_2$) и расположенных друг от друга на расстоянии, равном полусумме их фокусных расстояний, т. е.

$$l={}^1/_2(f_1+f_2) \tag{8}$$

Ахроматичность такой комбинации линз следует непосредственно из (5).

В приборе, состоящем из нескольких частей, в общем случае нельзя одновременно устранить хроматизм положения и хроматизм увеличения, если это не сделано для каждой его части. Докажем последнее утверждение для случая двух центрированных тонких линз, разнесенных на расстояние l. Согласно

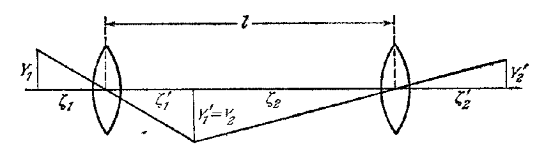


Рис. 4 26. Ахроматизация системы из двух тонких линз.

п. 4.4.4 отображение тонкой линзой является центральной проекцией из ее центра; следовательно (рис. 4.26),

$$\frac{Y_1'}{Y_1}=-\frac{\zeta_1'}{\zeta_1}, \qquad \frac{Y_2'}{Y_2}=-\frac{\zeta_2'}{\zeta_2}. \tag{9}$$

Поскольку $Y_2=Y_1'$, находим для увеличения

$$\frac{Y_2'}{Y_1}=\frac{\zeta_1'}{\zeta_1}\frac{\zeta_2'}{\zeta_2}. \tag{10}$$

Если длина волны изменится, то величина ζ_1 останется той же, величина ζ_2' также будет прежней, если допустить отсутствие хроматизма положения. Следовательно, условие отсутствия хроматизма увеличения системы можно записать в виде

$$\delta\left(\frac{\zeta_1'}{\zeta_2}\right)=\frac{1}{\zeta_2^2}(\zeta_2\delta\zeta_1'-\zeta_1'\delta\zeta_2)=0. \tag{11}$$

Так как $\zeta_2+\zeta_1'=l$, $\delta\zeta_1'=-\delta\zeta_2$, то (11) удовлетворяется лишь при $\delta\zeta_1'=\delta\zeta_2=0$, т. е. если каждая из этих линз ахроматизирована.

До сих пор мы рассматривали только первичную хроматическую аберрацию тонкой линзы и комбинации двух таких линз. В гл. 5 будут получены выражения для первичной хроматической аберрации центрированной системы в общем случае.

4.7.2. Дисперсия призмы. Рассмотрим теперь кратко прохождение света через призму.

Пусть α — угол между двумя рабочими гранями призмы. Предположим, что ребро A, вдоль которого пересекаются эти грани, перпендикулярно к плоскости, в которой лежат падающий, преломленный и вышедший лучи (рис. 4.27). Вначале будем считать свет строго монохроматическим.

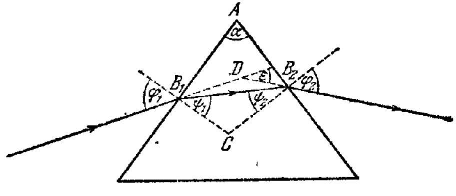


Рис. 4.27. Прохождение луча через призму.

Пусть B_1 и B_2 — точки пересечения падающего и вышедшего лучей с гранями призмы, φ_1 и ψ_1— углы падения и преломления в точке B_1 и ψ_2 и φ_2— внутренний и внешний углы в точке B_2 (т. е. углы, которые образуют луч B_1B_2 и вышедший луч с нормалью в точке B_2). Далее, пусть C — точка пересечения нормалей к граням призмы в B_1 и B_2 и D — точка, в которой пересекаются продолжения падающего и вышедшего лучей.

Если обозначить через ε *угол отклонения*, т. е. угол между вышедшим и падающим лучами, то

$$\varphi_1+\varphi_2=\varepsilon+\alpha, \tag{12}$$

$$\psi_1+\psi_2=\alpha. \tag{13}$$

Из закона преломления находим

$$\sin\varphi_1=n\sin\psi_1, \qquad \sin\varphi_2=n\sin\psi_2, \tag{14}$$

где n — показатель преломления стекла относительно окружающего воздуха. Угол отклонения ε экстремален, если

$$\frac{d\varepsilon}{d\varphi_1}=0. \tag{15}$$

С учетом (12) это условие принимает вид

$$\left(\frac{d\varphi_2}{d\varphi_1}\right)_{\text{экстр}}=-1. \tag{16}$$

Тогда из (13) и (14) найдем

$$\frac{d\psi_1}{d\varphi_1}=-\frac{d\psi_2}{d\varphi_2}, \qquad \cos\varphi_1=n\cos\psi_1\frac{d\psi_1}{d\varphi_1}, \qquad \cos\varphi_2\frac{d\varphi_2}{d\varphi_1}=n\cos\psi_2\frac{d\psi_2}{d\varphi_1}, \tag{17}$$

или после преобразований

$$\frac{d\varphi_2}{d\varphi_1}=-\frac{\cos\varphi_1\cos\psi_2}{\cos\psi_1\cos\varphi_2}. \tag{18}$$

Как следует из (16) и (18), в случае экстремума имеем

$$\frac{\cos\varphi_1\cos\psi_2}{\cos\psi_1\cos\varphi_2}=1, \tag{19}$$

или, возводя (19) в квадрат и используя (14), получим

$$\frac{1-\sin^2\varphi_1}{n^2-\sin^2\varphi_1}=\frac{1-\sin^2\varphi_2}{n^2-\sin^2\varphi_2}. \tag{20}$$

Это уравнение удовлетворяется, если

$$\left.\begin{aligned}&\varphi_1=\varphi_2;\\ \text{при этом}\quad &\psi_1=\psi_2.\end{aligned}\right\} \tag{21}$$

Чтобы определить характер экстремума, надо исследовать $d^2\varepsilon/d\varphi_1^2$. Из (12) и (18) имеем

$$\frac{d^2\varepsilon}{d\varphi_1^2}=\frac{d^2\varphi_2}{d\varphi_1^2}=\frac{d\varphi_2}{d\varphi_1}\frac{d}{d\varphi_1}\left[\ln\left(-\frac{d\varphi_2}{d\varphi_1}\right)\right]=$$

$$=\frac{d\varphi_2}{d\varphi_1}\left[-\operatorname{tg}\varphi_1-\operatorname{tg}\psi_2\frac{d\psi_2}{d\varphi_1}+\operatorname{tg}\psi_1\frac{d\psi_1}{d\varphi_1}+\operatorname{tg}\varphi_2\frac{d\varphi_2}{d\varphi_1}\right]. \tag{22}$$

При $\varphi_1 = \varphi_2$, $\psi_1 = \psi_2$ это выражение с учетом (14), (16) и (17) принимает вид

$$\left(\frac{d^2\varepsilon}{d\varphi_1^2}\right)_{\text{экстр}} = 2\,\mathrm{tg}\,\varphi_1 - 2\,\mathrm{tg}\,\psi_1 \frac{\cos\varphi_1}{n\cos\psi_1} = 2\,\mathrm{tg}\,\varphi_1\left(1 - \frac{\mathrm{tg}^2\psi_1}{\mathrm{tg}^2\varphi_1}\right). \tag{23}$$

Так как $n > 1$, то $\varphi_1 > \psi_1$; кроме того, поскольку $0 < \varphi_1 < \pi/2$, то $\mathrm{tg}\,\varphi_1 > 0$. Следовательно, $(d^2\varepsilon/d\varphi_1^2) > 0$, т. е. *отклонение минимально*. Из (21) следует, что при таком отклонении лучи проходят через призму *симметрично*. Величина угла наименьшего отклонения равна

$$\varepsilon_{\text{мин}} = 2\varphi_1 - \alpha. \tag{24}$$

Углы падения и преломления на первой грани призмы можно выразить через $\varepsilon_{\text{мин}}$ и α в виде

$$\varphi_1 = \frac{1}{2}(\varepsilon_{\text{мин}} + \alpha), \qquad \psi_1 = \frac{1}{2}\alpha, \tag{25}$$

откуда

$$n = \frac{\sin\varphi_1}{\sin\psi_1} = \frac{\sin\left[\frac{1}{2}(\varepsilon_{\text{мин}} + \alpha)\right]}{\sin\left(\frac{1}{2}\alpha\right)}. \tag{26}$$

Последняя формула часто используется при определении показателя преломления стекла. С помощью спектрометра измеряют значения $\varepsilon_{\text{мин}}$ и α, а затем по формуле (26) вычисляют величину n.

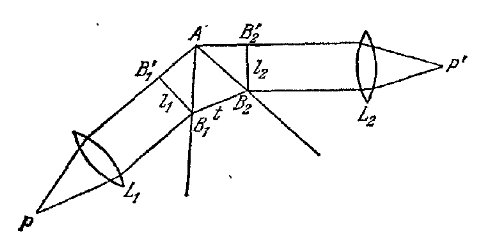


Рис. 4.28. Схема, иллюстрирующая дисперсию призмы

Рассмотрим теперь прохождение через призму пучка параллельных лучей, выходящих, например, из точечного источника P, расположенного в фокальной плоскости линзы L_1 (рис. 4.28); свет по-прежнему считается монохроматическим. Пусть B_1 и B_2' — основания перпендикуляров, опущенных из точек B_1 и B_2 на лучи, которые проходят через ребро A. Тогда B_1B_1' и B_2B_2' являются линиями пересечения двух волновых фронтов с плоскостью падения (плоскостью чертежа). Эти две линии образуют между собой угол, равный углу отклонения ε. Положим

$$B_1B_1' = l_1, \qquad B_2B_2' = l_2, \qquad B_1B_2 = t. \tag{27}$$

Рассмотрим теперь параллельный пучок немонохроматического света. Если линза L_1 исправлена на хроматическую аберрацию, то B_1B_1' останется на волновом фронте падающего пучка. С другой стороны, линия B_2B_2' уже не будет единственной; ее положение будет зависеть от длины волны λ, поскольку показатель преломления призмы зависит от длины волны, т. е.

$$n = n(\lambda), \tag{28}$$

а следовательно, угол отклонения ε тоже зависит от λ:

$$\varepsilon = \varepsilon(\lambda). \tag{29}$$

Величину

$$\frac{d\varepsilon}{d\lambda} = \frac{d\varepsilon}{dn}\frac{dn}{d\lambda}, \tag{30}$$

соответствующую постоянному значению угла падения φ_1, часто называют *угловой дисперсией призмы*. Первый сомножитель в правой части (30) зависит только от геометрии системы, а второй характеризует относительную дисперсию

стекла, из которого изготовлена призма. Поскольку $\varphi_1 = \text{const}$, то из (12) и (13) находим

$$\frac{d\varepsilon}{dn} = \frac{d\varphi_2}{dn}, \qquad \frac{d\psi_1}{dn} = -\frac{d\psi_2}{dn}, \tag{31}$$

а из (14) —

$$\sin\psi_1 + n\cos\psi_1 \frac{d\psi_1}{dn} = 0,$$

$$\cos\varphi_2 \frac{d\varphi_2}{dn} = \sin\psi_2 + n\cos\psi_2 \frac{d\psi_2}{dn} = 0; \tag{32}$$

после преобразований имеем

$$\frac{d\varepsilon}{dn} = \frac{\sin(\psi_1 + \psi_2)}{\cos\psi_2\cos\psi_1} = \frac{\sin\alpha}{\cos\varphi_2\cos\psi_1}. \tag{33}$$

Из треугольника $A_1B_1B_2$ получим на основании теоремы синусов

$$AB_2 = \frac{\cos\psi_1}{\sin\alpha}t, \tag{34}$$

а из треугольника AB_2B_2' —

$$AB_2 = l_2 \sec\varphi_2. \tag{35}$$

Используя последние три соотношения, можно представить (30) в виде

$$\frac{d\varepsilon}{d\lambda} = \frac{t}{l_2}\frac{dn}{d\lambda}. \tag{36}$$

Из симметрии следует, что в положении минимума отклонения $l_1 = l_2$. Если, кроме того, линзы настолько велики, что пучок полностью заполняет призму, то величина t будет равна длине b основания призмы. В этом случае из (36) получаем следующее выражение для *угловой дисперсии* $\delta\varepsilon$, т. е. для угла между волновыми фронтами, относящимися к длинам волн λ и $\lambda + \delta\lambda$:

$$\delta\varepsilon = \frac{b}{l_1}\frac{dn}{d\lambda}\delta\lambda. \tag{37}$$

§ 4.8. Фотометрия. Апертуры оптических систем

Фотометрией называется раздел оптики, связанный с измерениями световых потоков. Строго говоря, фотометрия не относится к геометрической оптике, однако во многих практических приложениях приближенная геометрическая картина электромагнитного поля служит при фотометрических исследованиях достаточно хорошей основой, и поэтому целесообразно включить в настоящую главу краткое рассмотрение этого раздела. Ограничимся простой геометрической моделью, согласно которой свет представляет собой поток лучистой энергии, распространяющийся вдоль геометрических лучей и подчиняющийся закону сохранения энергии. Последний состоит в том (см. уравнение (3.1.31)), что энергия, протекающая в единицу времени через любое поперечное сечение трубки лучей, остается постоянной.

4.8.1. Основные понятия фотометрии *). В фотометрии обычно рассматривается энергия света, испускаемого элементом поверхности S. Эта поверхность может быть либо фиктивной, либо реальной; в последнем случае она может совпадать с излучающей поверхностью источника или с освещенной поверхностью твердого тела. Если тело непрозрачно, то исследуется отраженный свет, если же оно прозрачно или полупрозрачно (в этом случае происходит частичное поглощение или рассеяние света), то измеряется обычно прошедший свет.

*) Здесь будут рассмотрены только фундаментальные фотометрические понятия. Более подробную информацию и описание приборов, используемых при световых измерениях, можно найти, например, в [39, 40].

Пусть $P(\xi, \eta)$ — произвольная точка на поверхности S, отнесенная к какой-то криволинейной системе координат на этой поверхности. Количество энергии (усредненное по времени), прошедшее за единицу времени через элемент поверхности δS, содержащий точку P, в телесный угол $\delta\Omega$ в направлении, определяемом полярными углами (α, β), можно записать следующим образом:

$$\delta F = B \cos\theta\, \delta S\, \delta\Omega. \tag{1}$$

Здесь θ — угол между направлением (α, β) и нормалью к элементу поверхности (рис. 4.29) и B — коэффициент, в общем случае зависящий от (ξ, η) и (α, β), т. е.

$$B = B(\xi, \eta; \alpha, \beta). \tag{2}$$

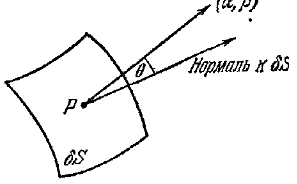


Рис. 4.29. Схема, иллюстрирующая распространение энергии от элемента поверхности

В выражении (1) стоит множитель $\cos\theta$, поскольку физический смысл имеет не сам элемент поверхности δS, а его проекция на плоскость, перпендикулярную к направлению (α, β). Величина B называется *фотометрической яркостью* в точке (ξ, η) в направлении (α, β). Необходимо отличать эту величину от субъективного ощущения яркости, поскольку глаз человека обладает разной чувствительностью по отношению к различным цветам *); этот вопрос будет обсуждаться более подробно в дальнейшем.

Величина δF записывается обычно двумя способами, чтобы показать в явном виде ее зависимость от $\delta\Omega$ и δS:

$$\delta F = \delta I\, \delta\Omega = \delta E\, \delta S. \tag{3}$$

Сравнивая (1) и (3), получим

$$\delta I = \frac{\delta F}{\delta\Omega} = B \cos\theta\, \delta S, \tag{4}$$

$$\delta E = \frac{\delta F}{\delta S} = B \cos\theta\, \delta S. \tag{5}$$

Интеграл

$$I(\alpha, \beta) = \int B \cos\theta\, dS, \tag{6}$$

взятый по определенной поверхности, называется *фотометрической силой света в направлении* (α, β), а интеграл

$$E(\xi, \eta) = \int B \cos\theta\, d\Omega, \tag{7}$$

взятый по телесному углу,— *фотометрической освещенностью в точке* **) (ξ, η).

Характер изменения B с направлением будет зависеть от свойств поверхности и особенно от того, шероховатая она или гладкая, обладает она способностью сама излучать свет или пропускает, или отражает падающее на нее излучение. Часто можно полагать, что B с хорошей точностью не зависит от направления. В этом случае говорят, что излучение *изотропно*. Если излучение изотропно и излучает плоская поверхность, то (6) принимает вид

$$I(\alpha, \beta) = I_0 \cos\theta, \tag{8}$$

где $I_0 = \int B\, dS$.

*) Мы используем прилагательное «фотометрический», когда хотим подчеркнуть, что данная величина определяется не по зрительному ощущению, а по физическому воздействию.

**) Если элемент поверхности δS в точке P перпендикулярен к вектору Пойнтинга, то фотометрическая освещенность δE в точке P совпадает с интенсивностью света, определенной в гл. 1.

Следовательно, фотометрическая сила света в каком-то направлении пропорциональна косинусу угла между этим направлением и нормалью к поверхности. Соотношение (8) называется *законом Ламберта (законом косинусов)*; если он выполняется, то говорят о *диффузном излучении*, или *диффузном отражении*, в соответствии с тем, имеем мы дело с излучающей или отражающей поверхностью.

Измерения величин F, B, I и E состоят в определении временного промежутка, площади, направления, телесного угла и энергии. Поскольку средние значения указанных величин обычно малы, необходимо использовать чувствительные приборы. Последние делятся на два типа. Первые реагируют на тепло, выделяющееся в поглощающей среде (например, болометр, термопара и др.), и используются главным образом при изучении теплового излучения (инфракрасного); вторые основаны на явлении поверхностного фотоэффекта, состоящего в том, что свет, падающий на поверхность металла, вызывает эмиссию электронов (если свет монохроматический, то число выбитых электронов (т. е. ток) пропорционально энергии падающего света). Приборы второго типа применяются, например, в фотографии в качестве экспонометров.

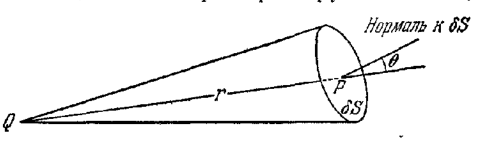


Рис. 4.30. К выводу закона освещенности в случае точечного источника.

Однако в технической фотометрии обычно используются косвенные методы. Сначала создают стандарт источника света и выражают его фотометрические показатели в абсолютных энергетических единицах. Затем результаты измерения сравнивают с этими показателями, причем часто в качестве нуль-индикатора, т. е. индикатора одинаковой яркости, используют глаз. Сравнение с эталоном основано на простом законе, справедливом для освещенности, создаваемой точечным источником *) Q.

Пусть δS — элемент поверхности в точке P, и пусть $QP = r$. Если θ — угол, образованный QP с нормалью к δS (рис. 4.30), то энергия, посылаемая источником через δS в единицу времени, равна $I\delta\Omega$, где I — фотометрическая сила света источника в направлении QP, а $\delta\Omega$ — телесный угол, под которым виден элемент δS из точки Q. Из элементарной геометрии следует, что

$$\cos\theta\,\delta S = r^2\,\delta\Omega. \tag{9}$$

Отсюда, используя (3), получим

$$E = \frac{I\cos\theta}{r^2}. \tag{10}$$

Соотношение (10) представляет собой основное уравнение практической фотометрии. Оно выражает *закон освещенности*, утверждающий, что освещенность E, создаваемая точечным источником, пропорциональна $\cos\theta$ и обратно пропорциональна r^2. Этот закон позволяет сравнивать силу света источников путем простых геометрических построений. Если элемент поверхности δS освещается двумя точечными источниками Q_1 и Q_2, фотометрические силы света которых равны I_1 и I_2, а линии, соединяющие δS с источниками, образуют с нор-

*) Фотометрическая сила света I точечного источника определяется с помощью процедуры, часто используемой при введении предельных величин. Предположим, что площадь поверхности стремится к нулю, а значение B — к бесконечности таким образом, что интеграл (6) остается конечным. Тогда I будет функцией α, β и положения точечного источника.

При расчете фотометрической освещенности обычно пренебрегают конечной протяженностью источника, если его линейные размеры меньше 1/20 его расстояния до освещаемой поверхности.

малью к δS углы θ_1 и θ_2 (рис. 4.31), то в случае равенства фотометрических освещенностей имеем

$$I_1 : I_2 = \frac{r_1^2}{\cos\theta_1} : \frac{r_2^2}{\cos\theta_2}. \qquad (11)$$

Сравнивая данный источник со стандартным, можно с помощью (11) определить его силу света. Равенство освещенностей, создаваемых двумя источниками, можно установить либо физическими методами, либо непосредственно глазом. Последний способ достаточно прост, если монохроматический свет от двух источников имеет одинаковую частоту, однако в общем случае приходится сравнивать источники, излучающие свет различного спектрального состава. Тогда

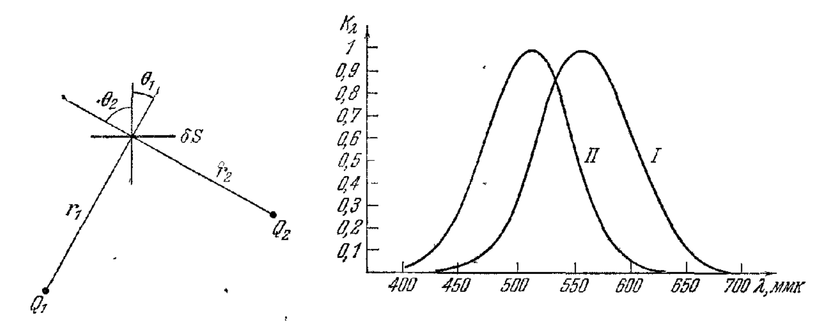


Рис. 4.31. К сравнению силы света двух точечных источников

Рис. 4.32 Кривая относительной спектральной чувствительности K_λ для среднего человеческого глаза:
I — для яркого света, *II* — для слабого света

этот простой метод уже не годится, поскольку глаз обладает различной чувствительностью к свету разных длин волн. В подобных случаях используют фильтры, пропускающие свет в узком спектральном интервале около известной длины волны; тогда установление равенства освещенностей для различных цветов сводится к определению относительных значений энергии с учетом *кривой спектральной чувствительности глаза.* Эта кривая характеризует зависимость от длины волны величины K_λ, обратной значению потока, вызывающего ощущение одинаковой яркости. Вид кривой до некоторой степени зависит от величины освещенности. Для яркого света она имеет максимум при λ около 550 *ммк* (см. рис. 4.32 и табл. 4.2). С уменьшением освещенности кривая сохраняет свою форму, но максимум смещается в сторону синего конца спектра; при очень слабом свете максимум лежит при λ, близком к 507 *ммк*. Это явление называется *эффектом Пуркинье.*

Если поток энергии оценивается по создаваемому им зрительному ощущению, а не по его истинной физической величине, то говорят о *световой энергии* F':

$$F' = \frac{\int K_\lambda F_\lambda \, d\lambda}{\int K_\lambda \, d\lambda}, \qquad (12)$$

где $F_\lambda d\lambda$ — количество энергии в интервале $(\lambda, \lambda + d\lambda)$, K_λ — относительная спектральная чувствительность, или относительная видность, и интегрирование проводится по всем длинам волн. Величины B', I' и E', связанные с B, I и E,

Таблица 4.2

Таблица значений относительной спектральной чувствительности K_λ среднего человеческого глаза (яркий свет)

λ, *ммк*	K_λ	λ, *ммк*	K_λ	λ, *ммк*	K_λ
400	0,0004	530	0,862	650	0,107
410	0,0012	540	0,954	660	0,061
420	0,0040	550	0,995	670	0,032
430	0,0116	560	0,995	680	0,017
440	0,023	570	0,952	690	0,0082
450	0,038	580	0,870	700	0,0041
460	0,060	590	0,757	710	0,0021
470	0,091	600	0,631	720	0,00105
480	0,139	610	0,503	730	0,00052
490	0,208	620	0,381	740	0,00025
500	0,323	630	0,265	750	0,00012
510	0,503	640	0,175	760	0,00006
520	0,710				

как F' с F, называются соответственно *яркостью, силой света и освещенностью* *). Эти понятия широко применяются в визуальной фотометрии.

Каждая из четырех величин F', B', I' и E' имеет свою единицу измерения в практической системе. Обычно основной фотометрической единицей считается единица силы света и через нее уже выражают единицы для F', B' и E', поскольку эталон силы света изготовить гораздо проще, чем эталон светового потока. Ранее за единицу силы света принималась *международная свеча*; соответствующие эталонные источники (угольные лампы) хранились в ряде национальных лабораторий. Недавно была введена новая единица, называемая *свечой (св)*; последняя определяется как одна шестидесятая силы света на квадратный сантиметр от абсолютно черного тела при температуре затвердевания платины ($\sim 2042°$ К). Значение силы света для излучения различного спектрального состава должно оцениваться с помощью описанной выше процедуры с учетом кривой спектральной чувствительности глаза.

Единица светового потока называется *люменом*. Она соответствует световому потоку, посылаемому внутрь единичного телесного угла однородным точечным источником с силой света в 1 свечу.

Величина единицы освещенности зависит от выбранной единицы длины. Метрической единицей освещенности является *люкс (лк)*, называемый иногда *метр-свеча*; она соответствует освещенности площади в 1 квадратный метр, на которую падает световой поток в 1 люмен.

Единицей яркости служат свеча на квадратный сантиметр, называемая *стильбом (сб)*, и свеча на квадратный метр — *нит*.

Иногда применяют и другие единицы, определения которых и их связь с рассмотренными здесь единицами можно найти в специальных книгах по фотометрии.

4.8.2. Диафрагмы и зрачки **). Количество света, достигающее пространство изображения оптической системы, зависит не только от яркости предмета, но и от размеров оптических элементов (линз, зеркал) и диафрагм. Диафрагма представляет собой отверстие, обычно круглое, в непрозрачном экране. Этот экран не пропускает лучи, претерпевшие сильную аберрацию. В этом разделе

---

*) В русской литературе термины «яркость», «сила света» и «освещенность» употребляются как для величин, связанных со световым восприятием B, I' и E', так и для соответствующих энергетических величин B, I и E. (*Прим. ред.*)

**) Теория диафрагмы была развита Аббе [41] и обобщена Рором [42].

под термином «диафрагмы» мы будем понимать не только отверстия в экране, но и края линз и зеркал.

Рассмотрим все лучи, исходящие из осевой точки предмета P_0. Диафрагма, определяющая поперечное сечение пучка, формирующего изображение, называется *апертурной диафрагмой*. Для определения ее положения необходимо найти параксиальное изображение каждой диафрагмы в части системы, предшествовавшей ей; изображение, которое видно из точки P_0 под наименьшим углом, называется *входным зрачком*. Реальная диафрагма, изображением которой является входной зрачок, является апертурной диафрагмой. (Если эта диафрагма находится перед первой поверхностью, то она совпадает с входным зрачком.) Угол $2\theta_0$, под которым виден диаметр входного зрачка из точки P_0, называется *угловой апертурой со стороны предмета* или просто *угловой апертурой* (рис. 4.33).

Изображение апертурной диафрагмы той частью системы, которая находится за этой диафрагмой (или изображение входного зрачка всей системой),

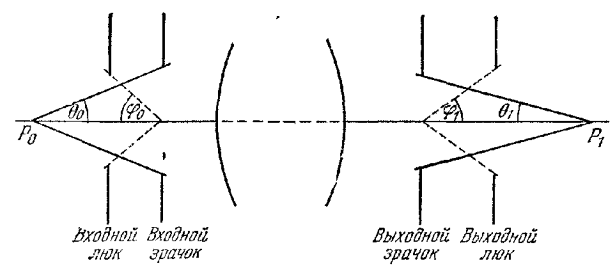


Рис. 4.33. Диафрагмы и зрачки.

называется *выходным зрачком*; угол $2\theta_1$, под которым виден его диаметр из точки P_1, можно назвать *угловой апертурой со стороны изображения* (или *углом проекции* системы).

В пучке лучей, исходящих из каждой точки предмета, всегда имеется луч, проходящий через центр входного зрачка; он называется *главным* (или *опорным*) *лучом* пучка и играет важную роль в теории аберраций. В отсутствие аберраций главный луч проходит также через центры апертурной диафрагмы и выходного зрачка.

Если апертурная диафрагма находится на задней фокальной плоскости той части системы, которая предшествует этой диафрагме, то входной зрачок расположен в бесконечности и все главные лучи в пространстве предмета будут параллельны оси. Такая система называется *телецентрической со стороны предмета*. Если апертурная диафрагма находится на передней фокальной плоскости той части системы, которая следует за этой диафрагмой, то выходной зрачок лежит в бесконечности, и все главные лучи в пространстве изображения будут параллельны оси; в этом случае система называется *телецентрической со стороны изображения*. Телецентрические системы используются при измерениях размеров предметов.

Если параметры системы, за исключением величины угловой апертуры, фиксированы, то последняя определяет количество света, проходящего через систему. Имеются и другие величины, часто используемые для описания «светосилы» оптической системы, например *числовая апертура* (Ч. А.) объектива микроскопа. Последняя определяется как произведение синуса от половины угловой апертуры со стороны предмета на показатель преломления среды в пространстве предмета, т. е.

$$\text{Ч.А.} = n \sin \theta_0. \tag{13}$$

Подходящей мерой, характеризующей светосилу систем, предназначенных для работы с предметами, расположенными на больших расстояниях, например телескопов или некоторых фотографических объективов, служит так называемое *«число F»* или *«номинальное фокальное отношение»* *). Оно равно отношению фокусного расстояния f объектива к диаметру входного зрачка d, т. е.

$$F = f/d. \qquad (14)$$

Так для линзы с фокусным расстоянием 10 *см* и апертурой 2 *см* величина $F = 5$. Такая линза называется линзой $f/5$.

Такие величины, как угловая апертура, числовая апертура и число F, которые можно считать мерой светосилы прибора, часто называют *относительными апертурами.*

Кроме апертурных диафрагм, в оптических системах существуют еще и *диафрагмы поля зрения*; они определяют ту часть поверхности протяженного предмета, которая отображается прибором. Рис. 4.34 иллюстрирует различие между этими двумя типами диафрагм.

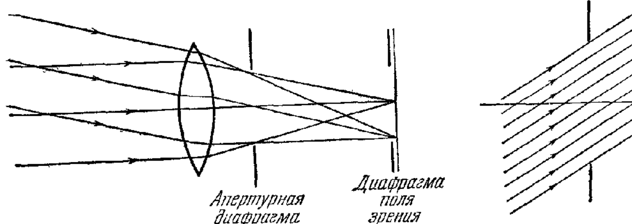


Рис. 4.34. Схема, иллюстрирующая различия между апертурной диафрагмой и диафрагмой поля зрения.

Рис. 4.35. Виньетирование.

Чтобы определить диафрагму поля зрения, найдем сначала изображение каждой диафрагмы той частью системы, которая предшествует ей. Изображение, которое видно из центра входного зрачка под наименьшим углом ($2\varphi_0$ на рис. 4.33), называется *входным люком*, а угол $2\varphi_0$— *полевым углом* или *углом поля зрения*. Диафрагма, которой соответствует входной люк, и является требуемой диафрагмой поля зрения.

Изображение входного люка всей системой (или изображение диафрагмы поля зрения той частью системы, которая следует за ней) называется *выходным люком*. Угол, под которым виден выходной люк из центра выходного зрачка ($2\varphi_1$ на рис. 4.33), иногда называется *угловым полем изображения.*

Может случиться, что некоторые лучи полностью минуют линзу или на часть ее не попадут лучи от ряда областей предмета, даже если величина апертурной диафрагмы меньше диаметра линзы. Такое срезание части пучка называется *виньетированием* (рис. 4.35). Подобное явление наблюдается довольно редко в системах, обладающих относительно малым полем зрения (например, в телескопах), однако оно играет важную роль в других приборах, например в фотообъективах. Иногда при конструировании оптических приборов используют эффект виньетирования для уничтожения нежелательных внеосевых аберраций.

4.8.3. Яркость и освещенность изображений. Рассмотрим теперь кратко соотношения между основными фотометрическими величинами, которые характеризуют излучение в пространствах изображения и предмета.

*) В русской литературе обычно используется обратная величина d/f, называемая «светосилой». *(Прим. ред.)*

Предположим, что предметом служит небольшой элемент плоскости δS_0, перпендикулярный к оси и излучающий по закону Ламберта. Тогда фотометрическая яркость B_0 не зависит от направления. Количество энергии δF_0, падающей в единицу времени на кольцевой элемент входного зрачка с центром на оси, равно

$$\delta F_0 = B_0 \cos\gamma_0\, \delta S_0\, \delta\Omega_0, \tag{15}$$

где

$$\delta\Omega_0 = 2\pi \sin\gamma_0\, \delta\gamma_0, \tag{16}$$

а γ_0 — угол, который образует луч, проходящий через кольцо, с осью. Следовательно, если, как и раньше, обозначить половину угловой апертуры со стороны предмета через θ_0, то полный поток энергии, проходящей в единицу времени через входной зрачок, будет определяться соотношением

$$F_0 = 2\pi B_0\, \delta S_0 \int\limits_0^{\theta_0} \sin\gamma_0 \cos\gamma_0\, d\gamma_0 = \pi B_0\, \delta S_0 \sin^2\theta_0. \tag{17}$$

Поток энергии F_1, выходящей из выходного зрачка, можно выразить аналогичным образом, а именно

$$F_1 = \pi B_1\, \delta S_1 \sin^2\theta_1. \tag{18}$$

Величина F_1 не может превышать величину F_0 и может лишь равняться ей, если в системе отсутствуют потери из-за отражения, поглощения или рассеяния; следовательно,

$$B_1 \sin^2\theta_1\, \delta S_1 \leqslant B_0 \sin^2\theta_0\, \delta S_0. \tag{19}$$

Отношение $\delta S_1/\delta S_0$ равно квадрату поперечного увеличения M:

$$\frac{\delta S_1}{\delta S_0} = M^2. \tag{20}$$

Если предположить, что система удовлетворяет условию синусов (4.5.6), то

$$\frac{n_0 \sin\theta_0}{n_1 \sin\theta_1} = M. \tag{21}$$

Подставляя (20) и (21) в (19), получим

$$B_1 \leqslant (n_1/n_0)^2 B_0. \tag{22}$$

Из (22), в частности, следует, что в случае, *когда показатели преломления в пространствах предмета и изображения равны между собой, фотометрическая яркость изображения не может превышать фотометрическую яркость предмета и равна ей, только если потери в системе пренебрежимо малы.*

Предполагая, что эти потери ничтожны, находим из (18) и (22)

$$F_1 = \pi (n_1/n_0)^2 B_0\, \delta S_1 \sin^2\theta_1, \tag{23}$$

т. е. фотометрическая освещенность $E_1 = F_1/\delta S_1$ на осевой точке изображения P_1 равна

$$E_1 = \pi (n_1/n_0)^2 B_0 \sin^2\theta_1. \tag{24}$$

При $\theta_1 \ll 1$ телесный угол Ω_1, под которым виден выходной зрачок из точки P_1, примерно равен $\pi \sin^2\theta_1$, так что (24) можно записать в виде

$$E_1 = (n_1/n_0)^2 B_0\, \Omega_1. \tag{25}$$

Эта формула получена для аксиального изображения; однако аналогичное выражение можно вывести и для неаксиального изображения. Если обозначить через φ_1 угол, образованный главным лучом CQ_1 с осью (рис. 4.36), то вместо (25) мы получим

$$E_1 = (n_1/n_0)^2 B_0\, \Omega_1' \cos\varphi_1, \tag{26}$$

где Ω_1' — телесный угол, под которым выходной зрачок виден из точки Q_1.

С помощью (9) можно написать

$$\Omega_1'/\Omega_1 = \cos\varphi_1\,(CP_1/CQ_1)^2 = \cos^3\varphi_1, \qquad (27)$$

т. е.

$$E_1 = (n_1/n_0)^2\, B_0\, \Omega_1 \cos^4\varphi_1. \qquad (28)$$

Полученное выражение говорит о том, что в случае, когда предмет излучает согласно закону Ламберта и отсутствуют потери в системе, а половина угловой апертуры θ_1 мала, *освещенность в некоторой точке изображения пропорциональна четвертой степени косинуса от угла, образованного главным лучом, проходящим через эту точку, с осью.*

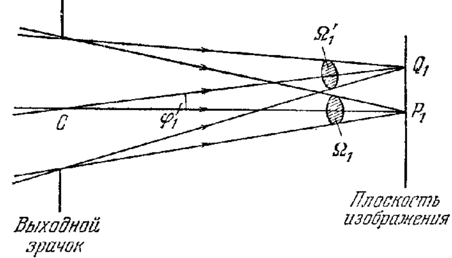


Рис. 4.36. Освещенность в точке внеаксиального изображения.

При использовании выведенных ранее формул необходимо помнить, что они были получены с помощью законов геометрической оптики. Эти формулы могут оказаться несправедливыми для очень маленьких источников. Например, изображением точечного источника является не точка, а яркий диск, окруженный кольцами (диск Эйри, см. п. 8.5.2); в этом случае свет распределен по всей дифракционной картине, и, следовательно, величина освещенности в геометрическом фокусе меньше той величины, которая получается из формулы (24).

§ 4.9. Метод построения хода лучей *)

При конструировании оптических приборов траектории световых лучей нужно определять с гораздо большей точностью, чем та, которую дает параксиальная оптика. Для этого следует воспользоваться алгебраическим анализом, учитывающим в разложении характеристической функции члены более высокого порядка малости (см. гл. 5). Другой способ, позволяющий с помощью элементарной геометрии более точно определять траектории световых лучей состоит в последовательном применении закона преломления (или отражения); этот метод, который будет сейчас кратко изложен, называется *методом построения хода лучей*; он находит широкое применение на практике.

4.9.1. Наклонные меридиональные лучи**). Рассмотрим сначала ход наклонного меридионального луча, т. е. луча, выходящего из точки, не лежащей на оси. Пусть A — полюс первой поверхности системы. Предположим, что это преломляющая сферическая поверхность радиуса r с центром в точке C, которая разделяет среды с показателями преломления n и n'. Падающий луч OP (рис. 4.37), лежащий в меридиональной плоскости, характеризуется углом U, образованным им с осью, и расстоянием $L = AB$ между полюсом A и точкой B, в которой данный луч пересекает ось. Пусть I — угол между падающим лучом и нормалью PC. Соответствующие величины, относящиеся к преломленному лучу, отмечены штрихами.

*) Для более полного ознакомления с методом построения хода лучей см., например, [43—45]. В работе [46] описан метод построения хода лучей с помощью электронных вычислительных машин, в работе [47] был предложен метод построения хода лучей, пересекающих несферические поверхности; см. также [48].

**) Здесь, в п. 4.9.2 и в § 4.10 будут использоваться общепринятые в данном случае обозначения и правило знаков. Правило знаков для углов отличается от декартова правила знаков, употребляемого в остальных разделах книги, и приводится к нему, если положить $U=-\gamma$, $U'=-\gamma'$.

Здесь будет использоваться следующее правило знаков: величины r, L и L' считаются положительными, если C, B и B' расположены справа от A (предполагается, что лучи падают слева). Углы U и U' положительны, если можно совместить ось с лучами PB и PB' путем ее поворота по часовой стрелке соответственно вокруг B и B' на угол, меньший 90°. Углы I и I' положительны, если можно совместить падающий и преломленный лучи с нормалью PC путем поворота по часовой стрелке вокруг точки падения P на угол, меньший 90°.

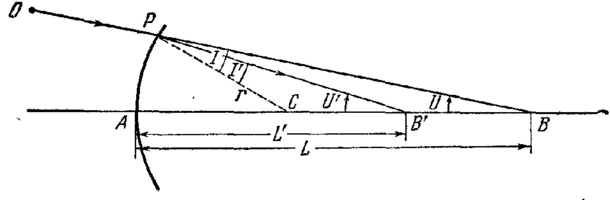


Рис. 4.37. Обозначения, применяемые при построении хода наклонного меридионального луча.

Предполагается, что величины L и U, характеризующие падающий луч, заданы и нужно вычислить L' и U'. Полагая также, что как L, так и r конечны, из треугольника PCB получим

$$\sin I = \frac{L-r}{r}\sin U. \tag{1}$$

Закон преломления дает

$$\sin I' = \frac{n}{n'}\sin I. \tag{2}$$

Из рисунка следует, что

$$U' = U + I - I'. \tag{3}$$

И, наконец, из треугольника PCB' находим

$$L' = \frac{\sin I'}{\sin U'}r + r. \tag{4}$$

Итак, последовательно применяя *законы преломления* (1) — (4), мы получили величины L' и U', характеризующие преломленный луч PB'.

Преломленный луч PB' является падающим относительно второй поверхности. Если написать L_1' вместо L', U_1' вместо U' и обозначить соответствующие величины через L_2, U_2 (L_2 отсчитывается от полюса второй поверхности), то мы получим *уравнения перехода*

$$L_2 = L_1' - d, \tag{5}$$

$$U_2 = U_1', \tag{6}$$

где $d > 0$ — расстояние между полюсами первой и второй поверхностей.

Подставим теперь величины (5) и (6) в уравнения (1) — (4). Решая эти уравнения относительно штрихованных величин, мы находим ход луча, прошедшего вторую поверхность. Многократно применяя законы преломления и уравнения перехода, мы получим значения величин L' и U', относящихся к лучу в пространстве изображения. После этого можно найти точку пересечения этого луча с плоскостью изображения. Практически, конечно, строят ход не одного, а нескольких, специально выбранных лучей; их точки пересечения с плоскостью изображения позволяют судить о работе данной системы.

Если одна из поверхностей (скажем, k-я) является зеркалом, то соответствующие формулы можно формально получить из уже выведенных, положив $n_k' = -n_k$. Тогда величину d_k следует считать отрицательной. Более того, все величины показателей преломления последующих сред и соответствующие значения d также нужно считать отрицательными до места второго отражения, когда они снова становятся положительными.

Рассмотрим теперь два частных случая, которые мы исключали до сих пор. Если падающий луч параллелен оси системы ($L=\infty$), то вместо уравнения (1) следует использовать соотношение

$$\sin I = Y/r, \tag{7}$$

где Y — расстояние от луча до этой оси (рис. 4.38, *а*).

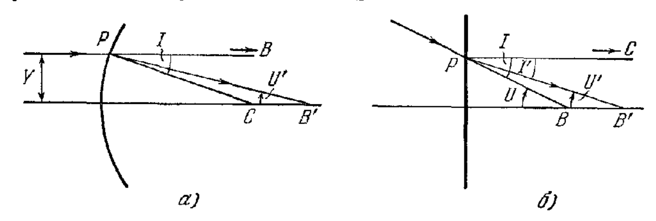


Рис. 4.38. Ход лучей в оптической системе.
а — падение лучей параллельно оси ($L=\infty$), *б* — падение лучей на плоскую поверхность ($r=\infty$).

В случае плоской поверхности ($r=\infty$) вместо (1)—(4) мы имеем следующую систему уравнений (рис. 4.38, *б*):

$$I = -U, \tag{8}$$

$$\sin U' = \frac{n}{n'} \sin U, \tag{9}$$

$$I' = -U', \tag{10}$$

$$L' = \frac{\operatorname{tg} U'}{\operatorname{tg} U} L. \tag{11}$$

Учитывая (9), уравнение (11) можно представить в виде

$$L' = \frac{n'}{n} \frac{\cos U'}{\cos U} L, \tag{11a}$$

который удобнее (11) при расчетах, если углы малы.

Полезно определить также координаты (Y_k, Z_k) точки падения P_k на k-ю поверхность и расстояние $D_k = P_k P_{k+1}$ между точками падения на соседние поверхности (рис. 4.39).

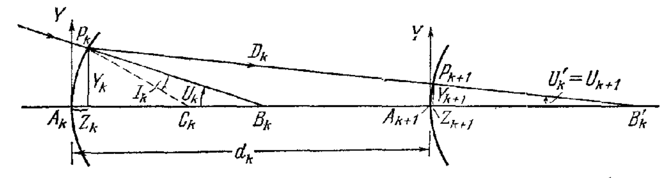


Рис 4.39. Ход наклонного меридионального луча, проходящего через две соседние преломляющие поверхности.

Из рисунка видно, что

$$Y_k = r_k \sin(U_k + I_k), \tag{12}$$

$$Z_k = r_k - r_k \cos(U_k + I_k) = \frac{Y_k^2}{[1+\cos(U_k+I_k)]\, r_k}, \tag{13}$$

$$D_k = \{d_k + Z_{k+1} - Z_k\} \sec U_k'. \tag{14}$$

Выразив L_k' через Y_k, Z_k и U_k', находим

$$L_k' = Z_k + \frac{Y_k}{\operatorname{tg} U_k'}. \tag{15}$$

Этим соотношением можно пользоваться для проверки величин, вычисленных с помощью (11а) или (11).

В частном случае, когда значение L_k равно бесконечности, $U_k = 0$, и соотношения (12)—(15) остаются справедливыми. Если $r_k = \infty$, то величину Y_k можно вычислить из выражения

$$Y_k = L_k \operatorname{tg} U_k, \tag{16}$$

а Z_k равно нулю.

4.9.2. Параксиальные лучи. Если угол между направлением луча и осью достаточно мал, то в полученных ранее формулах можно заменить синусы от различных углов на сами углы. Тогда эти формулы переходят в формулы параксиальной оптики. Последние используются на практике для вычисления гауссова увеличения и фокусного расстояния системы. Поэтому здесь будет дана краткая их сводка.

Принято обозначать величины, относящиеся к параксиальной области, маленькими буквами. Тогда законы преломления (1)—(4) запишутся следующим образом:

$$i = \frac{l-r}{r} u, \tag{17}$$

$$i' = \frac{n}{n'} i, \tag{18}$$

$$u' = u + i - i', \tag{19}$$

$$l' = \frac{i'}{u'} r + r. \tag{20}$$

Уравнения перехода (5) и (6) примут вид

$$l_2 = l_1' - d, \tag{21}$$

$$u_2 = u_1' \tag{22}$$

Из (7)—(11а) можно аналогичным образом получить соотношения параксиальной оптики для случаев $L = \infty$ и $r = \infty$.

Согласно (12) уравнение параксиальной оптики для высоты точки падения, которое потребуется в дальнейшем, имеет вид

$$y_k = r_k (u_k + i_k). \tag{23}$$

Хотя в соотношения (17)—(20) входят углы, образованные падающим и преломленным лучами с осью, величина l' не зависит от них. Этот результат, установленный в п. 4.4.1 другим способом, следует из настоящих формул, если из них исключить i, i' и u'. Оказывается, что величина u тоже исчезает, и окончательно получим

$$n' \left(\frac{1}{r} - \frac{1}{l'} \right) = n \left(\frac{1}{r} - \frac{1}{l} \right). \tag{24}$$

Это уравнение совпадает с соотношением Аббе (4.4.7). Для определения поперечного гауссового увеличения M необходимо построить лишь параксиальный луч, выходящий из осевой точки предмета. Тогда, согласно (4.4.54),

$$M = \frac{n_1}{n_l} \frac{u_1}{u_l}, \tag{25}$$

где индексы 1 и l относятся к первой и последней средам.

Фокусное расстояние системы f' можно получить, если построить параксиальный луч на любой заданной высоте y_1, исходящий от бесконечно удаленного предмета. Тогда уравнение сопряженного луча в пространстве изображения, отнесенное к системе координат с началом во втором фокусе, имеет вид $y_l/Z_l = -u_l'$, а из (4.3.10) следует, что

$$f' = -\frac{y_1}{u_l'}. \tag{26}$$

4.9.3. Косые лучи. До сих пор мы рассматривали только лучи, лежащие в меридиональной плоскости. Сейчас мы кратко исследуем прохождение *косых* лучей, т. е. лучей, не компланарных с осью. Построение траекторий таких лучей весьма трудоемко и обычно проводится лишь при конструировании систем с очень большими апертурами.

Будем теперь характеризовать луч направляющими косинусами и координатами точки, в которой он пересекает определенную поверхность системы. Выберем в полюсе A_1 первой поверхности прямоугольную систему координат с осью Z, направленной вдоль оси системы. Пусть L_1, M_1, N_1 ($L_1^2+M_1^2+N_1^2=1$) — направляющие косинусы луча, падающего в точку $P_1(X_1, Y_1, Z_1)$ (рис. 4.40).

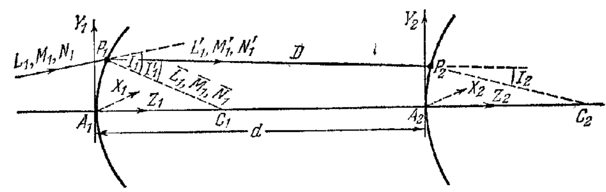


Рис. 4.40. Построение хода косого луча.

Первый шаг состоит в вычислении косинуса угла падения I_1. Если r_1 — радиус первой поверхности, то направляющие косинусы нормали в точке P_1 равны

$$\overline{L}_1=-\frac{X_1}{r_1},\qquad \overline{M}_1=-\frac{Y_1}{r_1},\qquad \overline{N}_1=\frac{r_1-Z_1}{r_1},$$

так что

$$\cos I_1=L_1\overline{L}_1+M_1\overline{M}_1+N_1\overline{N}_1=N_1-\frac{1}{r_1}(L_1X_1+M_1Y_1+N_1Z_1). \qquad (27)$$

Следующий шаг состоит в определении направляющих косинусов L_1', M_1', N_1' преломленного луча. Это делается в два приема. Сначала вычисляется косинус угла преломления с помощью закона преломления в форме

$$n'\cos I_1'=\sqrt{n'^2-n^2+n^2\cos^2 I_1}. \qquad (28)$$

Затем учитывается, что преломленный луч лежит в плоскости падающего луча и нормали к поверхности. Если обозначить через $\mathbf{s}_1$, $\mathbf{s}_1'$ и $\overline{\mathbf{s}}_1$ единичные векторы в направлениях падающего луча, преломленного луча и нормали к преломляющей поверхности, т. е. векторы с компонентами (L_1, M_1, N_1), (L_1', M_1', N_1') и ($\overline{L}_1$, $\overline{M}_1$, $\overline{N}_1$), то из условия компланарности получим

$$\mathbf{s}_1'=\lambda\mathbf{s}_1+\mu_1\overline{\mathbf{s}}_1, \qquad (29)$$

где λ и μ — некоторые скалярные функции. Для определения λ и μ умножим сначала (29) скалярно на $\mathbf{s}_1$ и используем соотношения (см. рис. 4.40) $\mathbf{s}_1\cdot\mathbf{s}_1'=\cos(I_1-I_1')$ и $\mathbf{s}\cdot\overline{\mathbf{s}}_1=\cos I_1$. Получим

$$\cos(I_1-I_1')=\lambda+\mu\cos I_1.$$

Затем, умножая (29) скалярно на $\overline{\mathbf{s}}_1$ и учитывая соотношения $\overline{\mathbf{s}}_1\cdot\mathbf{s}_1'=\cos I_1'$ и $\overline{\mathbf{s}}_1\cdot\mathbf{s}_1=\cos I_1$, находим

$$\cos I_1'=\lambda\cos I_1+\mu.$$

Из двух последних соотношений следует, что

$$\lambda=\frac{\sin I_1'}{\sin I_1}=\frac{n}{n'},\qquad \mu=\frac{1}{n'}(n'\cos I_1'-n\cos I_1),$$

и из (29) получаем три уравнения для направляющих косинусов преломленного луча, а именно

$$\left.\begin{aligned} n'L_1' &= nL_1 - \sigma X_1, \\ n'M_1' &= nM_1 - \sigma Y_1, \\ n'N_1' &= nN_1 - \sigma(Z_1 - r_1), \end{aligned}\right\} \tag{30}$$

где

$$\sigma = \frac{1}{r_1}(n' \cos I_1' - n \cos I_1). \tag{31}$$

Таким образом, с помощью уравнений (27), (28), (30) и (31) мы определили траекторию луча, прошедшего через первую поверхность.

Преломленный луч можно теперь считать падающим относительно второй поверхности. Выбирая систему координат с началом в полюсе второй поверхности A_2 и осями, параллельными осям первой системы, получим *уравнения перехода для направляющих косинусов* в виде

$$L_2 = L_1', \qquad M_2 = M_1', \qquad N_2 = N_1'. \tag{32}$$

Пусть d — расстояние A_1A_2 между полюсами первой и второй поверхностей. Обозначая через X_1^+, Y_1^+, Z_1^+ координаты точки P_1, отнесенные к системе с началом в точке A_2, получим *уравнения перехода для координат* в виде

$$X_1^+ = X_1, \qquad Y_1^+ = Y_1, \qquad Z_1^+ = Z_1 - d. \tag{33}$$

Далее необходимо определить координаты (X_2, Y_2, Z_2) точки P_2, в которой преломленный луч пересекает вторую поверхность. Если D — расстояние между P_1 и P_2, то

$$X_2 = X_1^+ + L_2D, \qquad Y_2 = Y_1^+ + M_2D, \qquad Z_2 = Z_1^+ + N_2D. \tag{34}$$

При определении D используется то обстоятельство, что P_2 лежит на второй поверхности. Пусть r_2— радиус этой поверхности; тогда

$$X_2^2 + Y_2^2 + Z_2^2 - 2Z_2r_2 = 0. \tag{35}$$

Подставляя (34) в это выражение, получим уравнение для D

$$D^2 - 2Fr_2D + Gr_2 = 0, \tag{36}$$

где

$$\left.\begin{aligned} F &= N_2 - \frac{1}{r_2}(L_2X_1^+ + M_2Y_1^+ + N_2Z_1^+), \\ G &= \frac{1}{r_2}[(X_1^+)^2 + (Y_1^+)^2 + (Z_1^+)^2] - 2Z_1^+. \end{aligned}\right\} \tag{37}$$

Уравнение (36) дает следующее значение D, если учесть, что для осевого луча $D = d$:

$$D = r_2\left(F - \sqrt{F^2 - \frac{1}{r_2}G}\right). \tag{38}$$

Подставляя полученное выражение для D в (34), находим координаты точки P_2.

И, наконец, для завершения цикла вычислений следует найти косинус угла падения I_2 на второй поверхности. Он определяется из выражения, совершенно аналогичного (27), а именно

$$\cos I_2 = N_2 - \frac{1}{r_2}(L_2X_2 + M_2Y_2 + N_2Z_2), \tag{39}$$

или с учетом (34) и (38) —

$$\cos I_2 = F - \frac{1}{r_2}D = \sqrt{F^2 - \frac{1}{r_2}G}. \tag{40}$$

Таким образом, последний шаг состоит в вычислении выражений (40), (38) и (34).

Из-за больших трудностей, возникающих при построении хода произвольных косых лучей, иногда ограничиваются случаем косых лучей, находящихся в непосредственной близости от выбранного главного луча. Построение хода таких лучей можно осуществить с помощью упрощенных методов (см. [49, 50]), сходных с методами, использованными при построении хода параксиальных лучей, и пригодных для определения положения сагиттальной фокальной поверхности.

§ 4.10. Оптические системы с несферическими поверхностями

В подавляющем большинстве оптических систем используются линзы и зеркала, поверхности которых имеют форму плоскости, сферы или параболоида. Выбор таких простых форм связан в основном с теми практическими трудностями, которые встречаются при изготовлении более сложных поверхностей с высокой степенью точности, необходимой в оптике. Использование поверхностей простой формы накладывает, естественно, ограничение на характеристики обычных оптических систем. Поэтому в некоторых системах применяют поверхности более сложной формы, называемые *асферическими*, несмотря на трудности, встречающиеся при их изготовлении. Еще в 1905 г. Шварцшильд [51] рассмотрел класс объективов телескопов *), состоящих из двух несферических зеркал, и показал, что такие системы можно сделать апланатическими.

В 1930 г. гамбургский оптик Бернард Шмидт сконструировал телескоп нового типа, состоящий из сферического зеркала и соответствующим образом рассчитанной асферической линзы, помещенной в центр кривизны зеркала. Оказалось, что такая система (она рассмотрена более подробно в § 6.4) обладает замечательными свойствами. С помощью этого телескопа удается сфотографировать на одной пластинке очень большой участок неба, в сотни раз превышающий по размерам участок, который можно сфотографировать при использовании телескопов обычной конструкции. С тех пор *камера Шмидта* стала очень важным инструментом при астрономических наблюдениях. Асферические системы, в которых используется принцип камеры Шмидта, применяются также в некоторых телевизионных приемниках проекторного типа (см., например, [52]), в рентгеновской фотографии с флуоресцирующим экраном и в некоторых скоростных спектрографах с низкой дисперсией. Асферические поверхности находят также полезное применение в микроскопии (см. § 6.6).

Если одну из поверхностей какой-либо центрированной системы сделать асферической, то в общем случае можно добиться полного осевого стигматизма; с помощью двух асферических поверхностей любую центрированную систему можно сделать в общем случае апланатической. В этом разделе будут выведены формулы, необходимые при конструировании таких поверхностей.

4.10.1. Получение осевого стигматизма **). Рассмотрим лучи, выходящие из осевой точки предмета P. В общем случае лучи из различных зон выходного зрачка пересекают ось в различных точках пространства изображений. Пусть $S^{(0)}$ — последняя поверхность системы и O — осевая точка этой поверхности (рис. 4.41). Покажем, что, подбирая форму поверхности $S^{(0)}$, можно полностью скомпенсировать отклонение пучка лучей, формирующих изображение, от гомоцентричности; иными словами, покажем, что, заменяя поверхность $S^{(0)}$ новой поверхностью S, можно в общем случае добиться пересечения всех лучей в пространстве изображения с осью в любой заданной точке Q.

*) Позднее были построены два телескопа такого типа: один с апертурой 60 *см* в Индианском университете и второй с апертурой 30 *см* в Университете Брауна.

**) Описанными здесь методами мы обязаны Вольфу и Предди [53, 54]. Аналогичные формулы были получены Лунебергом [23].

Из симметрии следует, что достаточно рассматривать только меридиональные лучи. Каждый луч, падающий на последнюю поверхность, характеризуется следующими параметрами: углом U, который он образует с осью, и расстоянием $H = OM$ от начала координат до точки пересечения луча с осью Y (см. рис. 4.41): удобно пометить все лучи. Пусть t — любой подходящий параметр, например угол U_0, образованный лучом с осью в пространстве предмета, или

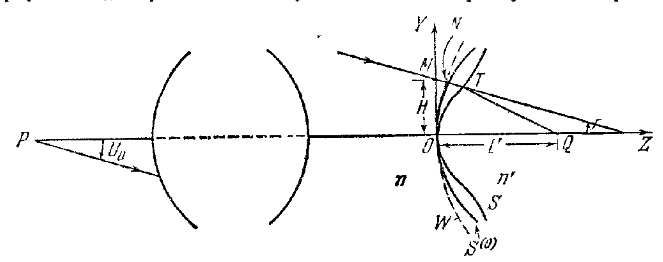


Рис 4.41. К расчету формы асферической поверхности, обеспечивающей осевой стигматизм.

высота точки пересечения луча с первой поверхностью. Тогда пучок полностью характеризуется двумя функциональными соотношениями

$$U = U(t), \qquad H = H(t). \tag{1}$$

Можно предположить, что $t = 0$ для осевого луча. В общем случае мы не знаем явного вида соотношений (1); однако методом построения хода лучей можно получить таблицу значений U и H для любого заданного набора значений t.

Пусть W — волновой фронт пучка, проходящий через точку O (см. рис. 4.41), а N—точка пересечения луча, идущего из точки M, с этой поверхностью. Предполагая, что можно найти скорректированную поверхность S с требуемыми свойствами, получим, что в исправленной таким образом системе оптическая длина пути от N до Q равна оптической длине пути от O до Q. Следовательно, если $T(Y, Z)$ — точка пересечения луча с S, то

$$[NT] + [TQ] = [OQ]. \tag{2}$$

Обозначая через n показатель преломления среды, расположенной перед поверхностью S, и через n'— показатель преломления среды в пространстве изображения, получим *)

$$\left.\begin{aligned} &[NT] = [MT] - [MN] = nZ \sec U - [MN], \\ &[TQ] = n' \sqrt{(L' - Z)^2 + (H - Z \operatorname{tg} U)^2}, \\ &[OQ] = n'L', \end{aligned}\right\} \tag{3}$$

где L'— расстояние от O до Q.

Оптическую длину пути $[MN]$, входящую в выражение для $[NT]$, также можно вычислить с помощью заданных величин. Применяя инвариантное соотношение Лагранжа (3.3.1) к линии, образованной отрезком OM оси Y, отрезком MN луча и кривой NO на волновом фронте W, получим

$$\int\limits_{OM} n\mathbf{s}\cdot d\mathbf{r} + \int\limits_{MN} n\mathbf{s}\cdot d\mathbf{r} + \int\limits_{NO} n\mathbf{s}\cdot d\mathbf{r} = 0, \tag{4}$$

где $\mathbf{s}$ — единичный вектор вдоль луча и $d\mathbf{r}$ — элемент пути интегрирования.

*) Как и в п. 4.9.1, считается, что угол U положителен, если можно совместить ось с лучом путем поворота по часовой стрелке вокруг осевой точки на угол, меньший 90°.

Далее, из рис. 4.41 следует, что

$$\left.\begin{aligned} &\int\limits_{OM} n\mathbf{s}\cdot d\mathbf{r} = -n\int\limits_0^H \sin U\, dH = -n\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt,\\ &\int\limits_{MN} n\mathbf{s}\cdot d\mathbf{r} = [MN], \qquad \int\limits_{NO} n\mathbf{s}\cdot d\mathbf{r} = 0; \end{aligned}\right\} \tag{5}$$

тогда (4) принимает вид

$$[MN] = n\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt. \tag{6}$$

Таким образом, величина $[MN]$ выражается через интеграл (6), который можно численно проинтегрировать с помощью таблицы значений U и H *).

Подставляя (3) и (6) в (2), получим следующее уравнение для Z:

$$\mathcal{A}Z^2 + 2\mathcal{B}Z\cos U + \mathcal{C}\cos^2 U = 0, \tag{7}$$

где

$$\left.\begin{aligned} &\mathcal{A} = n'^2 - n^2,\\ &\mathcal{B} = n^2\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt - n'^2(L'\cos U + H\sin U) + nn'L',\\ &\mathcal{C} = n'^2H^2 - \left(n\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt\right)\left(n\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt + 2n'L'\right). \end{aligned}\right\} \tag{8}$$

Следовательно,

$$Z = \frac{\cos U}{\mathcal{A}}\left[-\mathcal{B} \pm \sqrt{\mathcal{B}^2 - \mathcal{A}\mathcal{C}}\right]. \tag{9}$$

Из рис. 4.41 видно, что

$$Y = H - Z\operatorname{tg} U. \tag{10}$$

Перед корнем в выражении (9) необходимо взять положительный знак, так как $Z = U = H = 0$ при $t = 0$ и мы полагаем, что L' положительная величина (см. рис. 4.41). Наконец, комбинируя (9) и (10), получим

$$Z + iY = \frac{-\mathcal{B} + \sqrt{\mathcal{B}^2 - \mathcal{A}\mathcal{C}}}{\mathcal{A}}\,e^{-iU} + iH. \tag{11}$$

Последнее соотношение является точным параметрическим уравнением асферической поверхности S, выраженным через свободный параметр t.

Представляет интерес частный случай, когда фокус Q лежит на бесконечности ($L' = \infty$). При выводе соответствующих формул учтем, что величины $\mathcal{B}$ и $\mathcal{C}$ содержат L' только в первой степени. Следовательно, при достаточно больших L'

$$-\mathcal{B} + \sqrt{\mathcal{B}^2 - \mathcal{A}\mathcal{C}} = \\ = \mathcal{B}\left[-1 + \sqrt{1 - \frac{\mathcal{A}\mathcal{C}}{\mathcal{B}^2}}\right] = -\frac{1}{2}\frac{\mathcal{A}\mathcal{C}}{\mathcal{B}} - \frac{1}{8}\frac{\mathcal{A}^2\mathcal{C}^2}{\mathcal{B}^3} - \ldots = -\frac{1}{2}\frac{\mathcal{A}\mathcal{C}}{\mathcal{B}} + O\left(\frac{1}{L'}\right).$$

В пределе при $L' \to \infty$ выражение (11) сводится к

$$Z + iY = \frac{ne^{-iU}}{n - n'\cos U}\int\limits_0^t \sin U\,\frac{dH}{dt}\,dt + iH. \tag{12}$$

Мы рассмотрели только случай, когда асферическая поверхность является последней в системе. Однако использованный нами метод можно применить и к

*) Конечно, величину $[MN]$ можно непосредственно вычислить методом построения хода лучей, если воспользоваться тем, что оптическая длина пути от P до N равна оптической длине пути от P до O. Это дает $[MN]=[PO]-[PM]$.

расчету асферической поверхности, находящейся внутри оптической системы. В этих случаях вычисления значительно более громоздки, и поэтому мы их не будем рассматривать в настоящей книге *).

4.10.2. Получение апланатизма **). Мы показали, что, используя в системе одну асферическую поверхность, можно добиться точного осевого стигматизма. Рассмотрим теперь случай двух асферических поверхностей, позволяющих не только получить осевой стигматизм, но и обеспечить выполнение условия синусов.

Пусть S и S' — две асферические поверхности, форму которых необходимо определить. Предположим, что в данной системе между ними нет никаких поверхностей ***). Однако они могут быть отделены от точек предмета и изображения любым количеством преломляющих или отражающих поверхностей. Как и раньше, мы ограничимся только окончательной корректировкой системы,

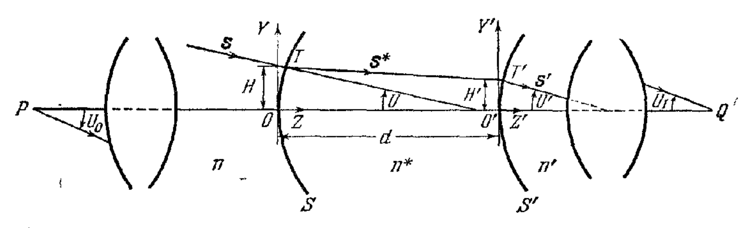


Рис. 4.42. К расчету формы двух асферических поверхностей системы, обеспечивающей апланатизм.

предполагая, что все расчетные параметры, за исключением профилей S и S', известны.

Выберем в полюсах O и O' поверхностей S и S' начала двух прямоугольных систем координат с осями Z, совпадающими с осью системы. Поверхность S будет рассматриваться относительно системы с началом в O, поверхность S' — относительно системы с началом в O'.

В пространстве, расположенном до поверхности S, пучок лучей, выходящий из осевой точки предмета P, характеризуется соотношением вида (рис. 4.42)

$$U = U(t), \qquad H = H(t). \tag{13}$$

Пучок лучей в пространстве за S' (в исправленной системе) характеризуется аналогичным соотношением

$$U' = U'(t'), \qquad H' = H'(t'). \tag{14}$$

Последние выражения можно протабулировать, строя ход лучей от выбранной осевой точки изображения P' в обратную сторону.

Если предмет и его изображение не лежат в бесконечности, то в качестве параметров t и t' выбирают синусы углов, которые образуют соответствующие лучи в пространствах предмета и изображения с осью системы ****):

$$t = \sin U_0, \qquad t' = \sin U_1. \tag{15}$$

Если предмет находится в бесконечности, то в качестве параметра t выбирают

*) Подробное исследование см. в работе [54]. Другие методы описаны в статьях [56 и 56].

**) Методы, описанные в настоящем разделе, разработаны Вассерманом и Вольфом [57].

***) Обобщение на случай систем, в которых S и S' не являются соседними, сделано в работе [58].

****) Если вместо условия синусов мы хотим (в дополнение к осевому стигматизму) удовлетворить условию Гершеля, то необходимо положить $t = \sin(U_0/2)$, $t' = \sin(U_1/2)$.

расстояние H_0 от соответствующего луча до оси в пространстве предмета; если же в бесконечности находится изображение, то $t' = H_1$, где H_1— расстояние от соответствующего луча до оси в пространстве изображения. В обоих этих случаях должно выполняться следующее соотношение, вытекающее из условия синусов:

$$t/t' = \text{const}. \tag{16}$$

Нашу задачу можно теперь сформулировать так: по данным (13) и (14) найти такие формы поверхностей S и S', которые обеспечили бы переход пучка (U, H) после двух последовательных преломлений на этих поверхностях в пучок (U', H'); кроме того, соответствующие лучи в двух пучках должны удовлетворять условию (16).

Пусть n — показатель преломления среды до поверхности S, n'— показатель преломления среды после S' и n^* — показатель преломления среды, находящейся между ними. Далее, пусть $\mathbf{s}$ — единичный вектор вдоль луча, падающего в точку $T(Z, Y)$, и $\mathbf{s}^*$— единичный вектор вдоль преломленного луча (см. рис. 4.42).

Согласно закону преломления (см. п. 3.2.2) вектор $\mathbf{N} = n\mathbf{s} - n\mathbf{s}^*$ должен быть параллелен нормали к поверхности S в точке T. Следовательно, если $\boldsymbol{\tau}$ — единичный вектор касательной в точке T меридионального сечения поверхности S, то

$$(n\mathbf{s} - n^*\mathbf{s}^*)\cdot\boldsymbol{\tau} = 0. \tag{17}$$

Таким образом, компоненты векторов $\mathbf{s}$, $\mathbf{s}^*$ и $\boldsymbol{\tau}$ будут равны

$$\left.\begin{array}{llll} \mathbf{s}: & 0, & -\sin U, & \cos U, \\ \mathbf{s}^*: & 0, & -\sin U^*, & \cos U^*, \\ \boldsymbol{\tau}: & 0, & \dfrac{\dot{Y}}{\sqrt{\dot{Y}^2+\dot{Z}^2}}, & \dfrac{\dot{Z}^2}{\sqrt{\dot{Y}^2+\dot{Z}^2}}, \end{array}\right\} \tag{18}$$

где U^* — угол, образованный преломленным лучом TT' с осью системы, а точка обозначает дифференцирование по параметру t. Уравнение (17) принимает вид

$$n(\dot{Z}\cos U - Y\sin U) = n^*(\dot{Z}\cos U^* - \dot{Y}\sin U^*). \tag{19}$$

Теперь, если обозначить через D расстояние между точками T и T', через D_y и D_z— проекции D на оси Y и Z и через d — расстояние вдоль оси между O и O', то имеем

$$\cos U^* = \frac{D_z}{D}, \qquad \sin U^* = \frac{D_y}{D}, \tag{20}$$

причем

$$D_y = Y - Y', \qquad D_z = d + Z' - Z, \qquad D = \sqrt{D_y^2 + D_z^2}. \tag{21}$$

Из рис. 4.42 следует, что

$$Y = H - Z\operatorname{tg} U, \tag{22}$$

$$Y' = H' - Z'\operatorname{tg} U'. \tag{23}$$

Подставляя в (19) значения $\cos U^*$ и $\sin U^*$ из (21) и значение Y из (22), получим

$$\frac{dZ}{dt} = \left(\frac{nD\cos U - n^*D_z}{nD\sin U - n^*D_y} + \operatorname{tg} U\right)^{-1}\left(\frac{dH}{dt} - Z\frac{d}{dt}\operatorname{tg} U\right). \tag{24}$$

Аналогично находим

$$\frac{dZ'}{dt} = \left(\frac{n'D\cos U' - n^*D_z}{n'D\sin U' - n^*D_y} + \operatorname{tg} U'\right)^{-1}\left(\frac{dH'}{dt'} - Z'\frac{d}{dt'}\operatorname{tg} U'\right). \tag{25}$$

Уравнения (21) — (25) вместе с соотношением (16) и граничными условиями $Z = Z' = 0$ при $t = t' = 0$ позволяют провести полный расчет скорректированных

профилей двух поверхностей. Так, с помощью (21) — (23) можно исключить Y и Y' из (24) и (25); далее, используя (16), получаем систему из двух дифференциальных уравнений первого порядка относительно Z и Z' вида

$$\frac{dZ}{dt}=f(Z,\ Z',\ t),\qquad \frac{dZ'}{dt}=g(Z,\ Z',\ t). \tag{26}$$

Эту систему можно решить обычными методами *). Однако, поскольку требуется определить не только Z и Z', но и Y и Y' для выбранных значений параметра t, проще не исключать из уравнений Y и Y', а последовательно вычислять неизвестные величины.

ЛИТЕРАТУРА

1. W. R. Hamilton, Trans. Roy. Irish Acad. **15**, 69 (1828); **16**, 1 (1830); **16**, 93 (1831), **17**, 1 (1837).
2. W. R. Hamilton, Mathematical Papers, Vol. I, Geometrical Optics, ed. by A. W. Conway a. J. L. Synge. Cambr Univ. Press, 1931
3. H. Bruns, Abh. Kgl. Sächs, Ges Wiss., math-phys. Kl. **21**, 323 (1895).
4. J. L. Synge, Geometrical Optics, Cambr. Univ. Press, 1937.
5. F. Klein, Z. Math. Phys 46, 376 (1901).
6. F. Klein, Ges. Math. Abh. **2**, 603 (1922).
7. C. Carathéodory, Geometrische Optik, Springer, Berlin, 1937, p. 4.
8. M. Herzberger, J. L. Synge, J. Opt. Soc. Amer. **27**, 75, 133, 138 (1937).
9. J. C. Maxwell, Quart. J. Pure Appl. Math. 2, 233 (1858).
10. J. C. Maxwell, Scientific Papers Vol. I, Cambr. Univ. Press. 1890.
11. H. Bruns, Das Ekonal, Abh Kgl. Sachs. Ges. Wiss., math.-phys. Kl. **21**, **370** (1895).
12. F. Klein, Ges. Math. Abh. 2, 607 (1922).
13. E. T. Whittaker, The Theory of Optical Instruments, Cambr. Univ. Press, 1907, p. 47.
14. H. Liebmann, Sitzgsber. Bayer. Akad. Wiss., Math. naturwiss. Abt. 183 (1916).
15. H. Boegehold, в сб. S. Czapski, O. Eppenstein, «Grundzuge der optischen Instrumente nach Abbe», Barth, Leipzig, 3rd ed., 1924, p. 213.
16. C. Carathéodory, Sitzgsber. Bayer. Akad. Wiss. Math.-naturwiss. Abt. **56**, 1 (1926).
17. J. Brown, Microwave Lenses, Methuen, London, 1953.
18. W. Lenz, в сб. «Probleme der Modernen Physik», ed by Debye, 1928, p. 198.
19. W. Blaschke, Vorlesungen über Differential-Geometrie, 4th ed. Springer, Berlin, 1945, p. 101.
20. F. Klein, Elementary Mathematics from an Advanced Standpoint, Macmillan, London, 1939, Vol. II (перевод с немецкого).
21. J. C. Maxwell, Cambr. a. Dublin Math. J. **8**, 188 (1854).
22. R. Stettler, Optik **12**, 529 (1955).
23. R. K. Luneberg, Mathematical Theory of Optics (мимеограф), Brown Univ., Providence, R. I., 1944, p. 213; печатное издание Univ. California Press, Berkeley a. Los Angeles, 1964, § 24.
24. R. F. Rinehart, J. Appl. Phys. **19**, 860 (1948).
25. A. Fletcher, T. Murphy, A. Young, Proc. Roy. Soc. **A223**, 216 (1954).
26. G. Toraldo di Francia, Optica Acta **1**, 157 (1954—1955).
27. H. Boegehold, M. Herzberger, Compositio Math. **1**, 448 (1935).
28. M. Herzberger, Ann. New York Acad. Sci. **48**, 1 (1946).
29. T. Smith, Proc. Phys. Soc. **60**, 293 (1948)
30. C. G. Wynne, Proc. Phys. Soc. **B65**, 436 (1952).
31. Rayleigh, Phil. Mag **21**, 466 (1886).
32. R. Clausius, Pogg Ann **121**, 1 (1864).
33. H. Helmholtz, Pogg. Ann Jubelband, 1874, p. 557.
34. W. Abbe, Jenaish. Ges Med. Naturwiss. 1879, p. 129.
35. E. Abbe, Carl. Repert Phys **16**, 303 (1880).
36. C. Hockin, J. Roy. Micro Soc. (2) **4**, 337 (1884).
37. J. F. W. Herschel, Phil. Trans. Roy. Soc. **111**, 226 (1821).
38. C. E. Weatherburn, Differential Geometry of Three Dimensions, Cambr. Univ. Press, 1927, p. 185.
39. J. W. T. Walsh, Photometry, Constable, 2nd ed., London, 1953.
40. Measurement of Radiant Energy, ed. by W. E. Forsythe, McGraw-Hill, N. Y., 1937.
41. E. Abbe, Jena Z. Naturwiss. 6, 263 (1871).
42. M. Rohr, Zentr. Z. Opt. Mech. **41**, 145, 159, 171 (1920).

*) Например, методом Адамса (см. [59]) или методом Рунге — Кутта (см. [60]).

43. A. E. Conrady, Applied Optics and Optical Design, Oxford Univ. Press, 1929, Pt. I.
44. M. Rohr, The Geometrical Investigation of Formation of Images in Optical Systems, Stationary Office, London, 1920 (перевод с немецкого).
45. M. Herzberger, Modern Geometrical Optics, Intersci. Publ., New York, 1958.
46. G. Black, Proc. Phys. Soc. **B68**, 569 (1954).
47. T. Smith, Proc. Phys. Sos **57**, 286 (1945).
48. W. Weinstein, Proc. Phys. Soc. **B65**, 731 (1952).
49. H. H. Hopkins, Proc. Phys. Soc. **58**, 663 (1946).
50. H. H. Hopkins, Wave Theory of Aberrations, Clarendon Press, Oxford 1950, pp. 59, 65.
51. K. Schwarzschild, Astr. Mitt. Königl. Sternwarte Göttingen, 1905 (перепечатано из Abh. Königl. Ges. Wiss. Gottingen, Math Phys. Kl. **4**, № 2 (1905—1906).
52. I. G. Maloff, D. W. Epstein, Electronics **17**, 98 (1944).
53. E. Wolf, W. S. Preddy, Proc. Phys. Soc. **59**, 704 (1947).
54. E. Wolf, Proc. Phys. Soc. **61**, 494 (1948).
55. M. Herzberger, H. O. Hoadley, J. Opt. Soc. Amer. **36**, 334 (1946).
56. D. S. Volosov, J. Opt. Soc. Amer. **37**, 342 (1947).
57. G. D. Wassermann, E. Wolf, Proc. Phys. Soc. **B62**, 2 (1949).
58. E. M. Vaskas, J. Opt. Soc. Amer. **47**, 669 (1957).
59. E. T. Whittaker, G. Robinson, The Calculus of Observations, Blackie a. Son. Glasgow, 4th, ed., 1946, p. 363.
60. C. Runge, H. König, Numerisches Rechnen, Springer, Berlin, 1924.

ГЛАВА 5

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ АБЕРРАЦИЙ

В § 4.9 было указано, что отклонение световых лучей от траекторий, предсказанных теорией Гаусса, можно исследовать в рамках геометрической оптики либо методом построения хода лучей, либо с помощью алгебраического анализа. В последнем методе, которому посвящена настоящая глава, при разложении характеристических функций оставляются члены, содержащие более высокие, чем вторая, степени расстояний от оси. Эти члены описывают *геометрические аберрации.*

Ранние попытки расширить теорию Гаусса связаны в основном с изобретением в 1839 г. Дагерром (1789—1851 гг.) фотографии. Перед практической оптикой, в которой до этого занимались главным образом конструированием объективов телескопов, встала новая задача создания объективов с большими апертурами и большими полями зрения. Венгерский математик Петцваль достиг большого успеха в решении этой проблемы, добавив в формулы Гаусса члены, содержащие более высокие степени углов наклона лучей относительно оси. К сожалению, его обширная монография, посвященная этому вопросу, была уничтожена ворами; все, что известно об его работе, содержится в полупопулярных статьях [1, 2]. Петцваль доказал практическую ценность своих вычислений, изготовив в 1840 г. широко известный портретный объектив (показанный на рис. 6.3, *б*), обладавший многими преимуществами по сравнению с известными тогда объективами. Наиболее раннее систематическое исследование геометрических аберраций, которое было полностью опубликовано, принадлежит Зайделю [3—5]; он учел все члены третьего порядка при рассмотрении общего случая центрированной системы, состоящей из сферических поверхностей. С тех пор это исследование было обобщено и упрощено многими авторами.

Поскольку волновые фронты ортогональны лучам, то при отклонении пучка лучей, формирующих изображение, от гомоцентричности форма соответствующих волновых фронтов отличается от сферической. Знание формы волновых фронтов играет существенную роль при более строгом исследовании аберраций, основанном на теории дифракции (см. гл. 9). По этой причине, а также из-за тесной связи между негомоцентричностью пучков и асферичностью соответствующих волновых фронтов мы будем рассматривать их совместно. Наше изложение будет частично основано на важной работе Шварцшильда [6] *), несколько упрощенной с этой целью.

§ 5.1. Волновые и лучевые аберрации; функция аберраций

Рассмотрим вращательно-симметричную оптическую систему. Пусть P_0', P_1' и P_1— точки пересечения луча, выходящего из точки предмета P_0, соответственно с плоскостью входного зрачка, плоскостью выходного зрачка и плоскостью параксиального изображения. Если P_1^*—параксиальное изображение точки P_0, то вектор $\delta_1 = \overrightarrow{P_1^* P_1}$ называется *аберрацией луча* или просто *лучевой аберрацией* (рис. 5.1).

*) Обобщение работы Шварцшильда на системы, не обладающие вращательной симметрией, проведено в статье [7].

Пусть W — волновой фронт, проходящий через центр O_1' выходного зрачка и связанный с пучком, который формирует изображение и выходит из точки P_0. Если аберрации отсутствуют, то W совпадает со сферой S, центр которой лежит

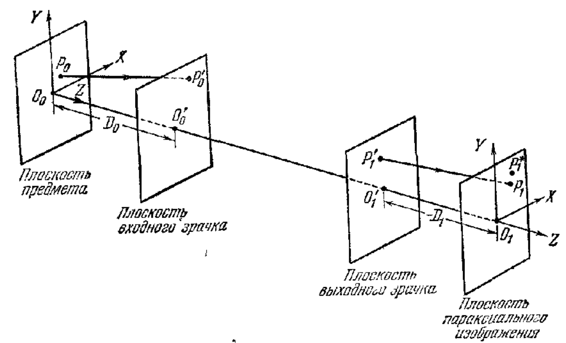


Рис. 5.1. Плоскость предмета, плоскость изображения и плоскости зрачков.

в точке параксиального изображения P_1^*, а сама она проходит через точку C_1'. S называется *опорной сферой Гаусса* (рис. 5.2).

Пусть Q и $\bar{Q}$ — точки пересечения луча $P_1'P_1$ с опорной сферой и волновым фронтом W соответственно. Оптическую длину пути $\Phi=[\bar{Q}Q]$ можно назвать *аберрацией волнового элемента в точке* Q или просто *волновой аберрацией* и считать положительной, если Q и P_1 расположены по разные стороны от $\bar{Q}$. В обычных приборах волновые аберрации достигают 40—50 длин волн, однако в приборах, используемых для более точных исследований (например, в астрономических телескопах или микроскопах), они должны быть значительно меньше, порядка долей длины волны.

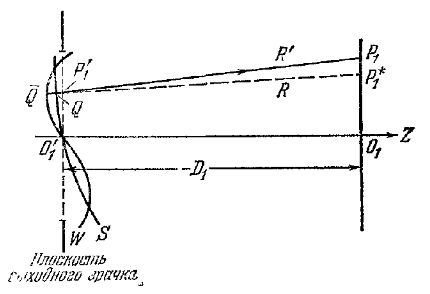


Рис. 5.2. Волновая и лучевая аберрации.

Выражения для волновой аберрации легко получить с помощью точечной характеристической функции Гамильтона системы.

Если, как и раньше, пользоваться для обозначения оптической длины пути квадратными скобками [...], то

$$\Phi=[\bar{Q}Q]=[P_0Q]-[P_0\bar{Q}]=[P_0Q]-[P_0O_1'].\qquad(1)$$

Здесь было использовано то обстоятельство, что точки $\bar{Q}$ и O_1' лежат на одном волновом фронте, т. е. $[P_0\bar{Q}]=[P_0O_1']$.

Введем две прямоугольные системы координат со взаимно параллельными осями, начала которых находятся в осевых точках O_0 и O_1 плоскостей предмета и изображения, а оси Z совпадают с осью системы. Точки в пространстве предмета будут рассматриваться в первой системе, а в пространстве изображения — во второй. Z-координаты плоскостей, в которых лежат зрачки, обозначены через D_0 и D_1 (на рис. 5.1 $D_1<0$).

Согласно (1) волновая аберрация выражается через точечную характеристику V следующим образом:

$$\Phi=V(X_0,\ Y_0,\ 0;\ X,\ Y,\ Z)-V(X_0,\ Y_0,\ 0;\ 0,\ 0,\ -D_1),\qquad(2)$$

где (X_0, Y_0) — координаты точки P_0 и (X, Y, Z) — координаты точки Q. Координаты (X, Y, Z) уже не являются независимыми; они связаны соотношением, учитывающим, что точка Q лежит на опорной сфере, т. е.

$$(X-X_1^*)^2+(Y-Y_1^*)^2+Z^2=R^2. \tag{3}$$

Здесь

$$X_1^*=MX_0, \qquad Y_1^*=MY_0 \tag{4}$$

— координаты точки P_1^* параксиального изображения, M — гауссово поперечное увеличение и R — радиус опорной сферы Гаусса

$$R=(X_1^{*2}+Y_1^{*2}+D_1^2)^{1/2}. \tag{5}$$

Величину Z в выражении (2) можно исключить с помощью (3), в результате чего Φ станет функцией только *) X_0, Y_0, X и Y, т. е.

$$\Phi=\Phi(X_0,\ Y_0;\ X,\ Y).$$

Лучевые аберрации связаны с функцией аберраций $\Phi(X_0, Y_0; X, Y)$ простыми соотношениями. Из (2) имеем

$$\frac{\partial\Phi}{\partial X}=\frac{\partial V}{\partial X}+\frac{\partial V}{\partial Z}\cdot\frac{\partial Z}{\partial X}. \tag{6}$$

Если α_1, β_1 и γ_1 — углы, которые образуют луч QP_1 с осями, а (X, Y, Z) и (X_1, Y_1, Z_1) — координаты точек Q и P_1, то, согласно (4.1.7) и рис. 5.2, получим

$$\frac{\partial V}{\partial X}=n_1\cos\alpha_1=n_1\frac{X_1-X}{R'}, \qquad \frac{\partial V}{\partial Z}=n_1\cos\gamma_1=-n_1\frac{Z}{R'}, \tag{7}$$

где

$$R'=\{(X_1-X)^2+(Y_1-Y)^2+Z^2\}^{1/2} \tag{8}$$

есть расстояние от Q до P_1 и n_1 — показатель преломления среды в пространстве изображения. Далее из (3) имеем

$$\frac{\partial Z}{\partial X}=-\frac{X-X_1^*}{Z}. \tag{9}$$

Подставляя (7) и (9) в соотношение (6), находим для компонент лучевой аберрации

$$X_1-X_1^*=\frac{R'}{n_1}\frac{\partial\Phi}{\partial X}$$

и аналогично

$$Y_1-Y_1^*=\frac{R'}{n_1}\frac{\partial\Phi}{\partial Y}. \tag{10}$$

Последние соотношения являются точными, но стоящая справа величина R' сама зависит от координат точки P_1, т. е. от лучевых аберраций **). Тем не менее для большинства практических целей R' можно заменять на радиус опорной сферы R или на другое приближенное выражение (см. ниже, уравнение (15)). Легко показать, что в силу симметрии задачи величина Φ зависит от четырех переменных, входящих только в трех комбинациях, а именно: $X_0^2+Y_0^2$, X^2+Y^2 и X_0X+Y_0Y. В самом деле, если ввести в плоскостях XY полярные координаты, т. е. положить

$$X_0=r_0\cos\theta_0, \quad X=r\cos\theta, \quad Y_0=r_0\sin\theta_0, \quad Y=r\sin\theta, \tag{11}$$

то окажется, что Φ зависит только от r_0, θ_0, r и θ, или, что то же самое, Φ зависит от r_0, r, $\theta_0-\theta$ и θ. Предположим теперь, что оси X и Y систем с началами в O_0 и O_1 поворачиваются на один и тот же угол и в одном и том же направ-

*) Вольф [8] ввел более общую функцию аберраций, удобную для исследования отображения протяженных предметов.

**) Несколько отличная пара точных уравнений, связывающих лучевые и волновые аберрации, была получена в [9].

лении относительно оси системы. При этом r_0, r и $\theta_0 - \theta$ не изменяются, а угол θ увеличивается на угол поворота. Поскольку функция Φ инвариантна относительно таких поворотов, она не должна зависеть от последней переменной, т. е. зависит только от r_0, r и $\theta_0 - \theta$. Следовательно, функция аберраций Φ является функцией трех скалярных произведений

$$\mathbf{r}_0^2 = X_0^2 + Y_0^2, \quad \mathbf{r}^2 = X^2 + Y^2 \quad \text{и} \quad \mathbf{r}_0 \cdot \mathbf{r} = X_0 X + Y_0 Y \tag{12}$$

двух векторов $\mathbf{r}_0(X_0, Y_0)$ и $\mathbf{r}(X, Y)$.

Отсюда вытекает, что при разложении Φ в ряд по степеням четырех координат нечетные степени будут отсутствовать. Поскольку $\Phi(0, 0; 0, 0) = 0$, то членов нулевой степени тоже не будет. Более того, не будет и членов второй степени, так как, согласно (10), они соответствуют лучевым аберрациям, линейно зависящим от координат, а это противоречит тому, что P_1^* является параксиальным изображением точки P_0. Таким образом, наше разложение имеет вид

$$\Phi = c\,(X_0^2 + Y_0^2) + \Phi^{(4)} + \Phi^{(6)} + \ldots \tag{13}$$

где c — константа, а $\Phi^{(2k)}$ — полином степени $2k$ по координатам и содержит их только в виде трех скалярных инвариантов (12). Говорят, что член степени $2k$ описывает *волновую аберрацию порядка* $2k$. Аберрации наинизшего порядка ($2k = 4$) обычно называются *первичными аберрациями* или *аберрациями Зайделя* *); они будут подробно рассмотрены в § 5.3.

Для оценки порядка величин некоторых выражений и точности наших вычислений удобно ввести параметр μ. Этим параметром может служить любая величина первого порядка, скажем, угловая апертура системы. Тогда можно допустить, что все лучи, проходящие через систему, составляют с оптической осью углы $O(\mu)$, где символ $O(\mu)$ означает, что величина угла порядка μ.

Оценим погрешность, возникающую при замене R' в основном уравнении (10) на величины, не зависящие от X_1 и Y_1. Из (3) и (5) имеем

$$Z^2 = D_1^2 - (X^2 + Y^2) + 2\,(XX_1^* + YY_1^*); \tag{14}$$

тогда вместо (8) можем написать

$$R' = -D_1\left[1 + \frac{X_1^2 + Y_1^2 - 2X(X_1 - X_1^*) - 2Y(Y_1 - Y_1^*)}{D_1}\right]^{1/2} =$$

$$= -D_1 - \frac{X_1^{*2} + Y_1^{*2}}{2D_1} + O(D_1\mu^4). \tag{15}$$

Соотношения (10) для компонент лучевой аберрации принимают вид

$$X_1 - X_1^* = -\frac{1}{n_1}\left(D_1 + \frac{X_1^{*2} + Y_1^{*2}}{2D_1}\right)\frac{\partial \Phi}{\partial X} + O(D_1\mu^7) = \tag{16а}$$

$$= -\frac{D_1}{n_1}\frac{\partial \Phi}{\partial X} + O(D_1\mu^5), \tag{16б}$$

$$Y_1 - Y_1^* = -\frac{1}{n_1}\left(D_1 + \frac{X_1^{*2} + Y_1^{*2}}{2D_1}\right)\frac{\partial \Phi}{\partial Y} + O(D_1\mu^7) = \tag{17а}$$

$$= -\frac{D_1}{n_1}\frac{\partial \Phi}{\partial Y} + O(D_1\mu^5). \tag{17б}$$

§ 5.2. Эйконал Шварцшильда

При исследовании геометрических аберраций Шварцшильд использовал метод, сходный с методом, применяемым в небесной механике при расчетах элементов орбит. В таких расчетах вводятся переменные, остающиеся постоянными при невозмущенном движении, а небольшие их изменения при действительном движении определяются с помощью функции возмущения. По анало-

*) Эти аберрации иногда называют *аберрациями третьего порядка*, так как связанные с ними *лучевые аберрации* содержат координаты в третьей степени.

гии с этим методом Шварцшильд в своих работах (ссылки на них уже приводились) ввел некоторые переменные, которые в приближении параксиальной оптики остаются постоянными вдоль каждого луча, проходящего через оптическую систему. Затем, введя некоторую функцию возмущения, названную им *эйконалом Зайделя*, он исследовал изменения этих переменных, возникающие при учете членов четвертого порядка в разложении характеристической функции *). Эти переменные были названы Шварцшильдом *переменными Зайделя*, поскольку они связаны с теми величинами, которые ранее использовал Зайдель.

Функция аберраций Φ, определенная выше, тесно связана в приближении теории Зайделя с возмущенным эйконалом Шварцшильда; используя метод Шварцшильда, можно получить выражения для коэффициентов четвертого порядка в разложении функции аберраций для произвольной центрированной системы. Это будет подробно проведено в § 5.5. Здесь мы определим переменные Зайделя и исследуем связь между нашей функцией аберраций и возмущенной функцией Шварцшильда.

Введем такие новые единицы длины в плоскостях предмета и параксиального изображения соответственно l_0 и l_1, чтобы

$$\frac{l_1}{l_0} = M, \tag{1}$$

где M — поперечное гауссово увеличение. Точки на плоскости предмета будем определять координатами x_0, y_0, а на плоскости изображения координатами x_1, y_1, для которых

$$x_0 = C\frac{X_0}{l_0}, \quad x_1 = C\frac{X_1}{l_1}, \quad y_0 = C\frac{Y_0}{l_0}, \quad y_1 = C\frac{Y_1}{l_1}, \tag{2}$$

где (X_0, Y_0), (X_1, Y_1) — обычные координаты точек P_0 и P_1 (см. рис. 5.1) и C — постоянная, которую мы определим позже. В приближении параксиальной оптики $x_1 = x_0$ и $y_1 = y_0$.

Координаты (X_0', Y_0') точек пересечения луча, исходящего из (X_0, Y_0), с входным зрачком связаны с лучевыми компонентами следующими соотношениями:

$$\frac{X_0' - X_0}{D_0} = \frac{p_0}{\sqrt{n_0^2 - p_0^2 - q_0^2}}, \quad \frac{Y_0' - Y_0}{D_0} = \frac{q_0}{\sqrt{n_0^2 - p_0^2 - q_0^2}}. \tag{3}$$

Соотношения для точек пересечения луча с выходным зрачком имеют совершенно аналогичный вид. Квадратные корни в знаменателях можно в приближении параксиальной оптики заменить соответственно на n_0 и n_1, в результате чего мы получим линейные соотношения между координатами вида

$$\left.\begin{aligned} \frac{X_0' - X_0}{D_0} &= \frac{p_0}{n_0}, & \frac{X_1' - X_1}{D_1} &= \frac{p_1}{n_1}, \\ \frac{Y_0' - Y_0}{D_0} &= \frac{q_0}{n_0}, & \frac{Y_1' - Y_1}{D_1} &= \frac{q_1}{n_1}. \end{aligned}\right\} \tag{4}$$

Теперь введем такие новые единицы длины λ_0 и λ_1 в плоскостях входного и выходного зрачков, чтобы

$$\lambda_1/\lambda_0 = M', \tag{5}$$

где M' — поперечное увеличение входной зрачок — выходной зрачок. Вместо

*) Шварцшильд исследовал также общий вид лучевых аберраций пятого порядка. Выражения для соответствующих коэффициентов шестого порядка в разложении характеристической функции (возмущенного эйконала Шварцшильда) были впервые получены его учеником Кольшюттером [10]. Аберрации пятого порядка рассматривались также в работах [11, 14] и других. Обзор теории аберраций более высоких порядков приведен в [12].

X'_0, Y'_0, X'_1, Y'_1 используем следующие переменные:

$$\left.\begin{aligned} \xi_0 &= \frac{X'_0}{\lambda_0} = \frac{X_0}{\lambda_0} + \frac{D_0 p_0}{\lambda_0 n_0}, & \xi_1 &= \frac{X'_1}{\lambda_1} = \frac{X_1}{\lambda_1} + \frac{D_1 p_1}{\lambda_1 n_1}, \\ \eta_0 &= \frac{Y'_0}{\lambda_0} = \frac{Y_0}{\lambda_0} + \frac{D_0 q_0}{\lambda_0 n_0}, & \eta_1 &= \frac{Y'_1}{\lambda_1} = \frac{Y_1}{\lambda_1} + \frac{D_1 q_1}{\lambda_1 n_1}. \end{aligned}\right\} \tag{6}$$

В приближении параксиальной оптики $\xi_1 = \xi_0$, $\eta_1 = \eta_0$.

Для упрощения дальнейших вычислений, удобно выбрать C в виде

$$C = \frac{n_0 l_0 \lambda_0}{D_0} = \frac{n_1 l_1 \lambda_1}{D_1}. \tag{7}$$

Другое выражение для C, часто используемое при изучении изображений с помощью теории дифракции, имеет вид

$$C = k n_0 l_0 \sin\theta_0 = -k n_1 l_1 \sin\theta_1.$$

Здесь k — волновое число для света в вакууме, а $2\theta_0$ и $2\theta_1$ — угловые апертуры системы, или углы, под которыми видны входной и выходной зрачки соответственно из осевых точек предмета и изображения. При таком значении C имеем

$$x_0 = k n_0 X_0 \sin\theta_0, \qquad x_1 = -k n_1 X_1 \sin\theta_1,$$
$$y_0 = k n_0 Y_0 \sin\theta_0, \qquad y_1 = -k n_1 Y_1 \sin\theta_1.$$

Равенство двух правых частей в выражении (7) следует из формулы Смита — Гельмгольца (4.4.48).

Величины, определенные соотношениями (2) и (6), являются *переменными Зайделя*. Нам потребуются и обратные соотношения, которые связывают старые переменные с переменными Зайделя. Решая (2) и (6), получим

$$\left.\begin{aligned} X_0 &= \frac{D_0}{n_0\lambda_0} x_0, & X_1 &= \frac{D_1}{n_1\lambda_1} x_1, \\ Y_0 &= \frac{D_0}{n_0\lambda_0} y_0, & Y_1 &= \frac{D_1}{n_1\lambda_1} y_1, \end{aligned}\right\} \tag{8}$$

$$\left.\begin{aligned} p_0 &= \frac{n_0\lambda_0}{D_0}\xi_0 - \frac{1}{\lambda_0} x_0, & p_1 &= \frac{n_1\lambda_1}{D_1}\xi_1 - \frac{1}{\lambda_1} x_1, \\ q_0 &= \frac{n_0\lambda_0}{D_0}\eta_0 - \frac{1}{\lambda_0} y_0, & q_1 &= \frac{n_1\lambda_1}{D_1}\eta_1 - \frac{1}{\lambda_1} y_1. \end{aligned}\right\} \tag{9}$$

Выразим теперь функцию аберраций через переменные Зайделя. Отметим сначала, что аргументы X и Y можно заменить в функции Φ на X'_1 и Y'_1, не изменяя членов $O(D_1\mu^5)$ в уравнениях (5.1.16б) и (5.1.17б). Обозначив функцию аберраций, выраженную через переменные Зайделя, через φ, имеем

$$\Phi(X_0, Y_0; X'_1, Y'_1) = \varphi(x_0, y_0; \xi_1, \eta_1). \tag{10}$$

Тогда

$$\frac{\partial\Phi}{\partial X'_1} = \frac{\partial\varphi}{\partial\xi_1}\frac{\partial\xi_1}{\partial X'_1} = \frac{1}{\lambda_1}\frac{\partial\varphi}{\partial\xi_1} \tag{11а}$$

и, учитывая (1), (2), (7) и (5.1.4),

$$X_1 - X_1^* = \frac{D_1}{n_1\lambda_1}(x_1 - x_0). \tag{11б}$$

С помощью уравнений (11) находим вместо (5.1.16б) и (5.1.17б) соотношения

$$x_1 - x_0 = -\frac{\partial\varphi}{\partial\xi_1} + O(D_1\mu^5), \qquad y_1 - y_0 = -\frac{\partial\varphi}{\partial\eta_1} + O(D_1\mu^5). \tag{12}$$

Ранее указывалось, что в приближении теории Зайделя φ тесно связана с функцией возмущений, введенной Шварцшильдом и названной им *эйконалом Зайделя*. Эта функция определяется соотношением

$$\psi = T + \frac{D_0}{2n_0\lambda_0^2}(x_0^2 + y_0^2) - \frac{D_1}{2n_1\lambda_1^2}(x_1^2 + y_1^2) + x_0(\xi_1 - \xi_0) + y_0(\eta_1 - \eta_0), \tag{13}$$

где $T = T(p_0, q_0, p_1, q_1)$ — угловая характеристика, отнесенная к системам с

началами координат в O_0 и O_1. Рассмотрим теперь небольшие вариации координат. Согласно уравнению (4.1.27) и условию $Z_0 = Z_1 = 0$, получим, что соответствующее изменение T равно

$$\delta T = X_0\, dp_0 + Y_0\, dq_0 - X_1\, dp_1 - Y_1\, dq_1, \tag{14}$$

или, выразив его через переменные Зайделя,—

$$\delta T = x_0\left(d\xi_1 - \frac{D_0}{n_0\lambda_0^2}\, dx_0\right) + y_0\left(d\eta_0 - \frac{D_0}{n_0\lambda_0^2}\, dy_0\right) - \\ - x_1\left(d\xi_1 - \frac{D_1}{n_1\lambda_1^2}\, dx_1\right) - y_1\left(d\eta_1 - \frac{D_1}{n_1\lambda_1^2}\, dy_1\right). \tag{15}$$

Используя последнее соотношение, находим из (13), что при малых изменениях переменных изменение ψ имеет вид

$$d\psi = (\xi_1 - \xi_0)\, dx_0 + (\eta_1 - \eta_0)\, dy_0 + (x_0 - x_1)\, d\xi_1 + (y_0 - y_1)\, d\eta_1, \tag{16}$$

т. е. ψ является функцией x_0, y_0, ξ_1, η_1, и поэтому получим строго

$$\left.\begin{aligned} \xi_1 - \xi_0 &= \frac{\partial\psi}{\partial x_0}, & x_1 - x_0 &= -\frac{\partial\psi}{\partial\xi_1}, \\ \eta_1 - \eta_0 &= \frac{\partial\psi}{\partial y_0}, & y_1 - y_0 &= -\frac{\partial\psi}{\partial\eta_1}. \end{aligned}\right\} \tag{17}$$

Следовательно, зная ψ, можно простым дифференцированием найти величины лучевых аберраций как в плоскости изображения, так и в плоскости выходного зрачка.

Сравнивая (17) и (12), видим, что в приближении теории Зайделя величина $\varphi - \psi$ не зависит от ξ_1 и η_1, т. е.

$$\varphi(x_0, y_0; \xi_1, \eta_1) = \psi(x_0, y_0; \xi_1, \eta_1) + \chi(x_0, y_0) + O(D_1\mu^6), \tag{18}$$

где $\chi(x_0, y_0)$ — некоторая функция от x_0 и y_0. Из определения φ следует, что $\varphi(x_0, y_0; 0, 0) = 0$; следовательно, $\chi(x_0, y_0) = -\psi(x_0, y_0; 0, 0)$ и

$$\varphi(x_0, y_0; \xi_1, \eta_1) = \psi(x_0, y_0; \xi_1, \eta_1) - \psi(x_0, y_0; 0, 0) + O(D_1\mu^6). \tag{19}$$

В области применимости теории Зайделя члены в (12), учитывающие погрешность, можно отбросить Однако если оставляются члены более высокого порядка чем четвертый, то соотношения для компонент лучевой аберрации, выраженные через функцию аберраций φ, принимают более сложный вид. Вместе с тем простые соотношения (17), выражающие компоненты лучевой аберрации через возмущенный эйконал, оказываются точными; однако возмущенный эйконал, по-видимому, не имеет простого физического смысла.

Определение членов более высокого порядка чем четвертый сопряжено, за исключением простейших случаев, с очень трудоемкими математическими выкладками. Поэтому обычно алгебраический анализ проводится в рамках теории Зайделя, которая затем в случае необходимости уточняется с помощью метода построения хода лучей.

§ 5.3. Первичные аберрации (аберрации Зайделя)

Используя рассуждения, совершенно аналогичные тем, которые относились в § 5.1 к функции аберраций, можно показать, что разложение в степенной ряд возмущенного эйконала Шварцшильда имеет в силу симметрии задачи следующий вид:

$$\psi = \psi^{(0)} + \psi^{(4)} + \psi^{(6)} + \psi^{(8)} + \ldots, \tag{1}$$

где $\psi^{(2k)}$ — полином степени $2k$ по четырем переменным; более того, эти переменные входят только в трех комбинациях:

$$r^2 = x_0^2 + y_0^2, \qquad \rho^2 = \xi_1^2 + \eta_1^2, \qquad \kappa^2 = x_0\xi_1 + y_0\eta_1. \tag{2}$$

В соотношении (1) отсутствует член второй степени, так как в противном случае это противоречило бы тому, что, согласно (5.2.17), $x_1 = x_0$, $y_1 = y_0$, $\xi_1 = \xi_0$ и $\eta_1 = \eta_0$ в приближении параксиальной оптики.

Поскольку переменные входят только в комбинациях (2), член $\psi^{(4)}$ должен иметь вид

$$\psi^{(4)} = -{}^1/_4 Ar^4 - {}^1/_4 B\rho^4 - C\kappa^4 - {}^1/_2 Dr^2\rho^2 + Er^2\kappa^2 + F\rho^2\kappa^2, \qquad (3)$$

где A, B, ... — постоянные. Знаки и числовые множители в (3) общепринятые; выражения для лучевых аберраций в этом случае принимают простой вид. Вычисление аберрационных постоянных для произвольной центрированной системы проводится в § 5.5.

Конечно, разложение в степенной ряд функции φ имеет такой же вид, как и (1), но оно не содержит члена нулевого порядка ($\varphi^{(0)} = 0$), и главный член $\varphi^{(4)}$ отличается от $\psi^{(4)}$ тем, что в нем отсутствует слагаемое $-\frac{1}{4}Ar^4$, как это непосредственно следует из уравнения (5.2.19). Таким образом, общее выражение для волновой аберрации наинизшего (четвертого) порядка записывается следующим образом:

$$\varphi^{(4)} = -\frac{1}{4}B\rho^4 - C\kappa^4 - \frac{1}{2}Dr^2\rho^2 + Fr^2\kappa^2 + F\rho^2\kappa^2, \qquad (4)$$

где B, C, ... — те же коэффициенты, что и в (3).

Подставляя последнее соотношение в уравнение (5.2.12), получим общее выражение для компонент лучевой аберрации наинизшего (третьего) порядка в виде

$$\left.\begin{aligned}
\Delta^{(3)}x &= x_1 - x_0 = \frac{n_1\lambda_1}{D_1}(X_1 - X_1^*) = \\
&= x_0(2C\kappa^2 - Er^2 - F\rho^2) + \xi_1(B\rho^2 + Dr^2 - 2F\kappa^2), \\
\Delta^{(3)}y &= y_1 - y_0 = \frac{n_1\lambda_1}{D_1}(Y_1 - Y_1^*) = \\
&= y_0(2C\kappa^2 - Er^2 - F\rho^2) + \eta_1(B\rho^2 + Dr^2 - 2F\kappa^2).
\end{aligned}\right\} \qquad (5)$$

Коэффициент A не входит в выражения (4) и (5), т. е. существуют только пять типов аберрации наинизшего порядка, характеризуемых пятью коэффициентами B, C, D, E и F. Как указывалось выше, эти аберрации называются *первичными* аберрациями или *аберрациями Зайделя*.

При исследовании аберраций Зайделя удобно выбрать оси таким образом, чтобы плоскость yz проходила через точку предмета; тогда $x_0 = 0$. Если затем ввести полярные координаты

$$\xi_1 = \rho \sin\theta, \qquad \eta_1 = \rho\cos\theta, \qquad (6)$$

то (4) примет вид

$$\varphi^{(4)} = -\frac{1}{4}B\rho^4 - Cy_0^4\rho^2\cos^2\theta - \frac{1}{2}Dy_0^2\rho^2 + Ey_0^3\rho\cos\theta + Fy_0\rho^3\cos\theta, \qquad (7)$$

а (5) — вид

$$\left.\begin{aligned}
\Delta^{(3)}x &= B\rho^3\sin\theta - 2Fy_0\rho^2\sin\theta\cos\theta + Dy_0^2\rho\sin\theta, \\
\Delta^{(3)}y &= B\rho^3\cos\theta - Fy_0\rho^2(1 + 2\cos^2\theta) + (2C + D)y_0^2\rho\cos\theta - Ey_0^3.
\end{aligned}\right\} \qquad (8)$$

В частном случае равенства нулю всех коэффициентов в (7) волновой фронт, проходящий через выходной зрачок совпадает (в рассматриваемом приближении) с опорной сферой Гаусса (см. рис. 5.2). В общем случае эти коэффициенты отличны от нуля. Тогда каждый член в (7) описывает определенный тип отклонения волнового фронта от правильной сферической формы; на рис. 5.3 показаны пять различных типов аберраций.

Важность лучевых аберраций, связанных с определенной точкой предмета, можно проиллюстрировать графически с помощью так называемых *аберрацион-*

ных (или *характеристических*) *кривых*. Эти кривые являются геометрическим местом точек пересечения лучей, выходящих из фиксированной зоны ρ = const выходного зрачка, с плоскостью изображения. Тогда поверхность, образованная аберрационными кривыми, соответствующими всем возможным значениям ρ, представляет собой неидеальное изображение.

Рассмотрим отдельно каждую из аберраций Зайделя *).

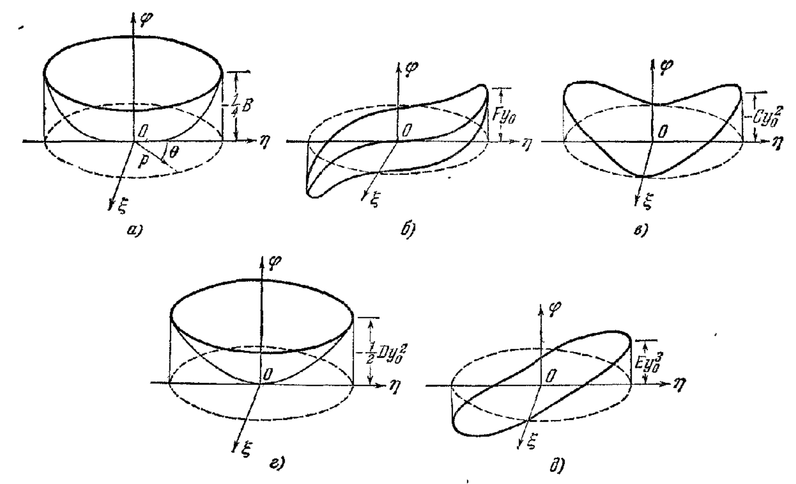


Рис. 5.3. Первичные волновые аберрации.

а — сферическая аберрация, $\varphi = -\frac{1}{4}B\rho^4$; *б* — кома, $\varphi = Fy_0\rho^3\cos\theta$; *в* — астигматизм: $\varphi = -Cy_0^2\rho^2\cos^2\theta$; *г* — кривизна поля, $\varphi = -\frac{1}{2}Dy_0^2\rho^2$; *д* — дисторсия, $\varphi = Ey_0^3\rho\cos\theta$.

I. *Сферическая аберрация* ($B \neq 0$). Если все коэффициенты, за исключением B, равны нулю, то (8) принимает вид

$$\Delta^{(3)}x = B\rho^3\sin\theta, \qquad \Delta^{(3)}y = B\rho^3\cos\theta. \tag{9}$$

Аберрационные кривые в этом случае имеют форму концентрических окружностей, центры которых расположены в точке параксиального изображения, а радиусы пропорциональны третьей степени радиуса зоны ρ, но не зависят от положения (y_0) предмета в поле зрения. Такой дефект изображения называется *сферической аберрацией*.

Сферическая аберрация, будучи независимой от y_0, искажает как осевые, так и внеосевые точки изображения. Лучи, выходящие из осевой точки предмета и составляющие существенные углы с осью, пересекут ее в точках, лежащих перед параксиальным фокусом или за ним (рис. 5.4). Точка, в которой пересекаются с осью лучи от края диафрагмы, называется *краевым фокусом*. Если экран в области изображения помещен под прямым углом к оси, то существует такое положение экрана, при котором круговое пятно изображения на нем минимально; это минимальное «изображение» называется *наименьшим кружком рассеяния*.

II. *Кома* ($F \neq 0$). Аберрация, характеризующаяся отличным от нуля коэффициентом F, называется комой. Компоненты лучевой аберрации в этом

*) Аберрационные кривые, соответствующие аберрациям более высокого порядка, рассматриваются в работах [13, 14].

случае имеют, согласно (8), вид

$$\left.\begin{aligned} \Delta^{(3)}x &= -2F\rho^2 y_0 \sin\theta\cos\theta = -Fy_0\rho^2 \sin 2\theta, \\ \Delta^{(3)}y &= -F\rho^2 y_0 (1+2\cos^2\theta) = -Fy_0\rho^2 (2+\cos 2\theta). \end{aligned}\right\} \qquad (10)$$

Как мы видим, при фиксированных y_0 и радиусе зоны ρ точка P_1 (см. рис. 5.1) при изменении θ от 0 до 2π дважды описывает в плоскости изображения окружность. Радиус окружности равен $|Fy_0\rho^2|$, а ее центр находится на расстоянии $2F\rho^2 y_0$ от параксиального фокуса в сторону отрицательных значений y. Следовательно, эта окружность касается двух прямых, проходящих через параксиальное изображение P_1 и составляющих с осью y углы в 30°. Если ρ пробегает все возможные значения, то совокупность подобных окружностей образует область, ограниченную отрезками этих прямых и дугой наибольшей аберрационной окружности (рис. 5.5). Размеры получающейся области линейно

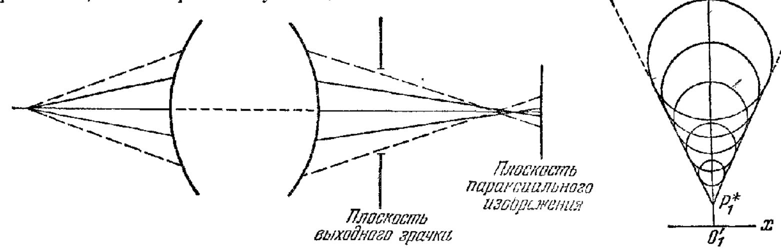


Рис. 5.4. Сферическая аберрация

Рис. 5.5. Кома.

возрастают с увеличением расстояния точки предмета от оси системы. В § 4.5 было показано, что при выполнении условия синусов Аббе система дает резкое изображение элемента плоскости предмета, расположенного в непосредственной близости от оси. Следовательно, в этом случае разложение функции аберрации не может содержать члены, линейно зависящие от y_0. Отсюда вытекает, что если условие синусов выполняется, первичная кома отсутствует.

III. *Астигматизм* ($C \neq 0$) *и кривизна поля* ($D \neq 0$). Аберрации, характеризующиеся коэффициентами C и D, удобнее рассматривать совместно. Если все остальные коэффициенты в (8) равны нулю, то

$$\Delta^{(3)}x = D\rho y_0^2 \sin\theta, \qquad \Delta^{(3)}y = (2C+D)\rho y_0^2 \cos\theta. \qquad (11)$$

Чтобы продемонстрировать важность таких аберраций, предположим вначале, что пучок, формирующий изображение, очень узок. Согласно § 4.6 лучи такого пучка пересекают два коротких отрезка кривых, одна из которых (тангенциальная фокальная линия) ортогональна меридиональной плоскости, а другая (сагиттальная фокальная линия) лежит в этой плоскости. Рассмотрим теперь свет, исходящий от всех точек конечной области плоскости предмета. Фокальные линии в пространстве изображения перейдут в *тангенциальную* и *сагиттальную фокальные поверхности*. В первом приближении эти поверхности можно считать сферами. Пусть R_t и R_s — их радиусы, которые считаются положительными, если соответствующие центры кривизны расположены по ту сторону от плоскости изображения, откуда распространяется свет (в случае, изображенном на рис. 5.6, $R_t > 0$ и $R_s < 0$).

Радиусы кривизны можно выразить через коэффициенты C и D. Для этого при вычислении лучевых аберраций с учетом кривизны удобнее использовать обычные координаты, а не переменные Зайделя. Имеем (рис. 5.7)

$$\frac{\Delta^{(3)}Y_1}{Y_1'} = \frac{u}{D_1+u}, \qquad (12)$$

где u — малое по величине расстояние между сагиттальной фокальной линией и плоскостью изображения. Если v — расстояние от этой фокальной линии до оси, то

$$R_t^2 = v^2 + (R_t - u)^2,$$

или

$$v^2 = 2R_t u - u^2.$$

Если считать u величиной первого порядка малости, то v можно заменить на

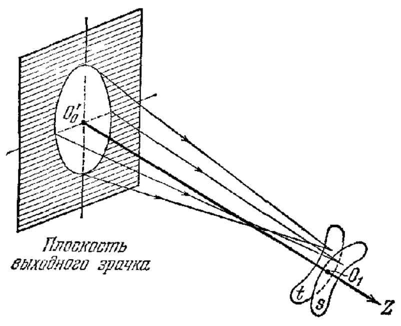


Рис 5.6. Тангенциальная и сагиттальная фокальные поверхности.

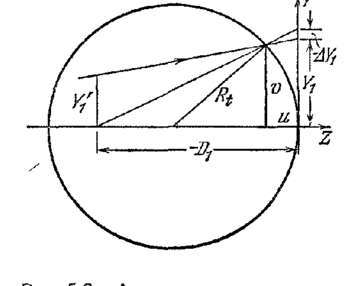


Рис. 5.7. Астигматизм и кривизна поля

Y_1, а в последнем уравнении отбросить u^2; тогда

$$u \approx \frac{Y_1^2}{2R_t}; \tag{13}$$

если еще пренебречь u по сравнению с D_1, то из (12) находим

$$\Delta^{(3)} Y_1 = \frac{Y_1^2}{2R_t} \frac{Y_1'}{D_1}. \tag{14}$$

Аналогично

$$\Delta^{(3)} X_1 = \frac{Y_1^2}{2R_s} \frac{X_1'}{D_1}. \tag{15}$$

Запишем теперь эти соотношения через переменные Зайделя. Подставляя в них (5.2.6) и (5.2.8), получим

$$\Delta^{(3)} y = \frac{n_1 \lambda_1}{D_1} \Delta^{(3)} Y_1 = \frac{n_1 \lambda_1}{D_1} \frac{Y_1^2}{2R_t} \frac{Y_1'}{D_1} = \frac{n_1 \lambda_1}{2D_1^2 R_t} \frac{D_1^2}{n_1^2 \lambda_1^2} y_1^2 \lambda_1 \eta_1,$$

или

$$\Delta^{(3)} y = \frac{y_1^2 \eta_1}{2n_1 R_t}, \tag{16}$$

и аналогично

$$\Delta^{(3)} x = \frac{y_1^2 \xi_1}{2n_1 R_s}. \tag{17}$$

В последних двух соотношениях y_1 можно заменить на y_0 и тогда, используя (11) и (6), получим

$$\frac{1}{R_t} = 2n_1 (2C + D), \qquad \frac{1}{R_s} = 2n_1 D. \tag{18}$$

Величину $2C + D$ обычно называют *тангенциальной кривизной поля*, величину D — *сагиттальной кривизной поля*, а их полусумму

$$\frac{1}{R} = \frac{1}{2}\left(\frac{1}{R_t} + \frac{1}{R_s}\right) = 2n_1 (C + D), \tag{19}$$

которая пропорциональна их среднему арифметическому значению,— просто *кривизной поля.*

Из (13) и (18) следует, что на высоте Y_1 от оси расстояние между двумя фокальными поверхностями (т. е. астигматическая разность пучка, формирующего изображение) равно

$$\frac{Y_1^2}{2}\left(\frac{1}{R_t}-\frac{1}{R_s}\right)=2n_1CY_1^2. \qquad (20)$$

Полуразность

$$\frac{1}{2}\left(\frac{1}{R_t}-\frac{1}{R_s}\right)=2n_1C \qquad (21)$$

называется *астигматизмом.* В отсутствие астигматизма ($C = 0$) имеем $R_t = R_s = R$. В п. 5.5.3 будет показано, что радиус R общей, совпадающей, фокальной поверхности можно в этом случае вычислить с помощью простой формулы, в которую входят радиусы кривизны отдельных поверхностей системы и показатели преломления всех сред.

IV. *Дисторсия* ($E \neq 0$). Если в соотношениях (8) отличен от нуля лишь коэффициент E, то

$$\Delta^{(3)}x=0, \qquad \Delta^{(3)}y=-Ey_0^3. \qquad (22)$$

Поскольку сюда не входят координаты ρ и θ, отображение получится стигматическим и не будет зависеть от радиуса выходного зрачка; однако расстояния точек изображения до оси не будут пропорциональны соответствующим расстояниям для точек предмета. Эта аберрация называется *дисторсией.*

При наличии тако аберрации изображение любой прямой в плоскости предмета, проходящей через ось, будет прямой линией, но изображение любой другой прямой будет искривленным. На рис. 5.8, *а* показан предмет в виде

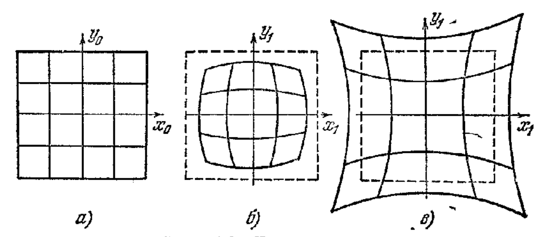


Рис. 5.8. Дисторсия.

а—предмет; *б*—бочкообразная дисторсия ($E > 0$), *в*—подушкообразная дисторсия ($E < 0$).

сетки прямых, параллельных осям x и y и расположенных на одинаковом расстоянии друг от друга. Рис. 5.8, *б* иллюстрирует так называемую *бочкообразную дисторсию* ($E > 0$), а рис. 5.8, *в* — *подушкообразную дисторсию* ($E < 0$).

Ранее указывалось, что из пяти аберраций Зайделя три (сферическая, кома и астигматизм) нарушают резкость изображения. Две другие (кривизна поля и дисторсия) изменяют его положение и форму. В общем случае невозможно сконструировать систему, свободную как от всех первичных аберраций, так и от аберраций более высокого порядка; поэтому всегда приходится искать какое-то подходящее компромиссное решение, учитывающее их относительные величины. В некоторых случаях аберрации Зайделя можно существенно уменьшить за счет аберраций более высокого порядка. В других случаях необходимо полностью уничтожить некоторые аберрации, несмотря на то, что при этом появляются аберрации других типов. Например, в телескопах должна быть полностью устранена кома, потому что при наличии ее изображение будет несимметричным и все прецизионные астрономические измерения положения поте-

ряют смысл. С другой стороны, наличие некоторой кривизны поля и дисторсии относительно безвредно, поскольку от них можно избавиться с помощью соответствующих вычислений.

До сих пор мы изучали аберрации только в рамках геометрической оптики. Однако если аберрации малы (волновые аберрации порядка длины волны или меньше), то становятся существенными дифракционные эффекты. В этом случае геометрическая теория должна быть дополнена более тонкими исследованиями, что и будет сделано в гл. 9.

§ 5.4. Теорема сложения для случая первичных аберраций

Продемонстрировав важность первичных аберраций, мы должны приступить к решению гораздо более трудной задачи, состоящей в вычислении коэффициентов первичных аберраций для случая произвольной центрированной системы. Как было показано, это эквивалентно определению членов четвертого порядка в разложении возмущенного эйконала Шварцшильда. Чтобы не прерывать основных вычислений, удобно вначале рассмотреть зависимость возмущенного эйконала системы от возмущенных эйконалов, связанных с отдельными поверхностями системы.

Рассмотрим центрированную систему, состоящую из двух поверхностей вращения, и пусть O_0 — осевая точка предмета, а O_1 и O_2 — ее параксиальные изображения, создаваемые соответственно первой и обеими поверхностями. Пусть далее

$$T_1 = T_1(p_0,\ q_0;\ p_1,\ q_1) \tag{1}$$

— угловая характеристика для преломления первой поверхностью, а

$$T_2 = T_2(p_1,\ q_1;\ p_2,\ q_2) \tag{2}$$

— угловая характеристика для преломления второй поверхностью, причем T_1 рассматривается относительно координатных систем с началами в O_0 и O_1, а T_2 — относительно координатных систем с началами в O_1 и O_2; оси X и Y этих систем взаимно параллельны, а оси Z направлены вдоль оси системы. Поскольку среда считается однородной, угловая характеристика представляет собой оптическую длину пути между основаниями перпендикуляров, опущенных из обоих начал координат на начальный и конечный отрезки луча (см. рис. 4.3). Следовательно, угловая характеристика системы T, отнесенная к системам с началами в O_0 и O_2, равна сумме T_1 и T_2, т. е.

$$T = T_1 + T_2. \tag{3}$$

Переменные p_1, q_1 в этом выражении, относящиеся к лучу в промежуточном пространстве, нужно исключить, пользуясь формулами, которые описывают отображения каждой из поверхностей. Это удобно сделать в явном виде в соответствующих выражениях для возмущенного эйконала.

Согласно (5.2.13), возмущенные эйконалы ψ_1 и ψ_2 имеют вид

$$\psi_1 = T_1 + \frac{D_0}{2n_0\lambda_0^2}(x_0^2 + y_0^2) - \frac{D_1}{2n_1\lambda_1^2}(x_1^2 + y_1^2) + x_0(\xi_1 - \xi_0) + y_0(\eta_1 - \eta_0)$$

и

$$\psi_2 = T_2 + \frac{D_1}{2n_1\lambda_1^2}(x_1^2 + y_1^2) - \frac{D_2}{2n_2\lambda_2^2}(x_2^2 + y_2^2) + x_1(\xi_2 - \xi_1) + y_1(\eta_2 - \eta_1),$$

а возмущенный эйконал ψ системы — вид

$$\psi = T + \frac{D_0}{2n_0\lambda_0^2}(x_0^2 + y_0^2) - \frac{D_2}{2n_2\lambda_2^2}(x_2^2 + y_2^2) + x_0(\xi_2 - \xi_0) + y_0(\eta_2 - \eta_0).$$

Следовательно, на основании (3) имеем

$$\psi = \psi_1 + \psi_2 + (x_0 - x_1)(\xi_2 - \xi_1) + (y_0 - y_1)(\eta_2 - \eta_1),$$

или, используя (5.2.17),

$$\psi = \psi_1 + \psi_2 + \frac{\partial \psi_1}{\partial \xi_1}\frac{\partial \psi_2}{\partial x_1} + \frac{\partial \psi_1}{\partial \eta_1}\frac{\partial \psi_2}{\partial y_1}. \tag{4}$$

Подставив в (4) разложения в степенные ряды величин ψ_1 и ψ_2, получим (напомним, что, согласно (5.3.1), члены второго порядка отсутствуют)

$$\psi = \psi_1^{(0)} + \psi_2^{(0)} + \psi_1^{(4)} + \psi_2^{(4)} + \ldots, \tag{5}$$

где точки означают члены шестого и более высокого порядков. Очевидно, что в выражениях для $\psi_1^{(4)}$ и $\psi_2^{(4)}$ мы имеем право заменить аргументы их гауссовыми значениями. Таким образом, переменные, относящиеся к промежуточному пространству, можно исключить с помощью следующего правила: заменить ξ_1, η_1 на ξ_2, η_2 в $\psi_1^{(4)}$ $(x_0, y_0; \xi_1, \eta_1)$ и x_1, y_1 на x_0, y_0 в $\psi_2^{(4)}$ $(x_1, y_1; \xi_2, \eta_2)$ и полученные выражения сложить. В результате получим $\psi^{(4)}$ как функцию четырех переменных x_0, y_0 и ξ_2, η_2, что и требовалось. Проведем теперь это исключение более подробно.

Согласно (5.3.3) $\psi_1^{(4)}$ и $\psi_2^{(4)}$ имеют вид

$$\left.\begin{aligned} \psi_1^{(4)} &= -\frac{1}{4}A_1 r_0^4 - \frac{1}{4}B_1\rho_1^4 - C_1\kappa_{01}^4 - \frac{1}{2}D_1 r_0^2\rho_1^2 + E_1 r_0^2\kappa_{01}^2 + F_1\rho_1^2\kappa_{01}^2, \\ &\text{где} \\ r_0^2 &= x_0^2 + y_0^2, \quad \rho_1^2 = \xi_1^2 + \eta_1^2, \quad \kappa_{01}^2 = x_0\xi_1 + y_0\eta_1, \end{aligned}\right\} \tag{6}$$

$$\left.\begin{aligned} \psi_2^{(4)} &= -\frac{1}{4}A_2 r_1^4 - \frac{1}{4}B_2\rho_2^4 - C_2\kappa_{12}^4 - \frac{1}{2}D_2 r_1^2\rho_2^2 + E r_1^2\kappa_{12}^2 + F_2\rho_2^2\kappa_{12}^2, \\ &\text{где} \\ r_1^2 &= x_1^2 + y_1^2, \quad \rho_2^2 = \xi_2^2 + \eta_2^2, \quad \kappa_{12}^2 = x_1\xi_2 + y_1\eta_2. \end{aligned}\right\} \tag{7}$$

Заменяя ξ_1, η_1 на ξ_2, η_2 и x_1, y_1 на x_0, y_0, т. е. ρ_1 на ρ_2, r_1 на r_0 и k_{12} на

$$\kappa_{02} = x_0\xi_2 + y_0\eta_2, \tag{8}$$

находим после сложения

$$\psi^{(4)} = -\frac{1}{4}(A_1 + A_2) r_0^4 - \frac{1}{4}(B_1 + B_2)\rho_2^4 - (C_1 + C_2)\kappa_{02}^4 - \\ -\frac{1}{2}(D_1 + D_2) r_0^2\rho_2^2 + (E_1 + E_2) r_0^2\kappa_{02}^2 + (F_1 + F_2)\kappa_{02}^2. \tag{9}$$

Сходный результат, конечно, получится и для произвольной центрированной системы, состоящей из любого числа поверхностей. Таким образом, мы доказали следующую теорему.

Каждый коэффициент первичной аберрации для любой центрированной системы равен сумме соответствующих коэффициентов для отдельных ее поверхностей.

Именно здесь особенно отчетливо видна практическая полезность переменных Зайделя, поскольку этот простой, но важный результат можно получить при использовании указанных переменных и он не имеет аналога для обычных переменных.

§ 5.5. Коэффициенты первичных аберраций произвольной центрированной системы линз

Сформулированная в предыдущем параграфе теорема позволяет свести задачу о нахождении коэффициентов первичных аберраций произвольной центрированной системы к вычислению соответствующих коэффициентов для отдельных ее поверхностей. Проведем теперь эти вычисления.

5.5.1. Формулы Зайделя, выраженные через параметры двух параксиальных лучей. Напомним, что коэффициенты аберраций Зайделя равны (с точностью до постоянных множителей) коэффициентам при членах четвертого порядка в разложении возмущенного эйконала Шварцшильда ψ. Согласно (5.2.13) эта функция получается, если к угловой характеристике T прибавить некоторые квадратичные члены, а результирующее выражение записать в переменных Зайделя. Поскольку переменные Зайделя и лучевые компоненты связаны линейными соотношениями, порядок членов при такой замене переменных не изменится. Следовательно,

$$\psi^{(4)}(x_0,\ y_0;\ \xi_1,\ \eta_1) = T^{(4)}(p_0,\ q_0;\ p_1,\ q_1). \tag{1}$$

Разложение угловой характеристики для преломляющей поверхности вращения до членов четвертого порядка малости включительно было проведено в § 4.1. Согласно (4.1.42) члены четвертого порядка можно записать в виде *)

$$T^{(4)}(p_0,\ q_0;\ p_1,\ q_1) = -\frac{r}{4(n_1-n_0)^2}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]\left[\frac{p_0^2+q_0^2}{n_0}-\frac{p_1^2+q_1^2}{n_1}\right] - \\ -\frac{(1+b)\,r}{8(n_1-n_0)^3}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2+\frac{a_0}{8n_0^3}(p_0^2+q_0^2)^2-\frac{a_1}{8n_1^3}(p_1^2+q_1^2)^2. \tag{2}$$

Удобно расположить осевые точки предмета и их параксиального изображения в $z=a_0$ и $z=a_1$ и положить (рис. 5.9)

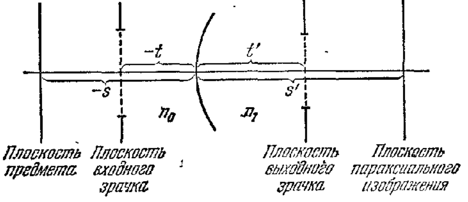


Рис. 5.9. К вычислению коэффициентов первичной аберрации.

$$s=a_0,\qquad s'=a_1,\qquad t=a_0+D_0,\qquad t'=a_1+D_1. \tag{3}$$

Обозначим через K и L соответствующие инварианты Аббе (см. (4.4.7))

$$n_0(1/r-1/s)=n_1(1/r-1/s')=K, \tag{4}$$

$$n_0(1/r-1/t)=n_1(1/r-1/t')=L. \tag{5}$$

Прежде чем подставлять выражения для лучевых компонент, записанные в переменных Зайделя, в (1), удобно переписать это соотношение в несколько иной форме. Согласно (4) имеем

$$\frac{r}{8(n_1-n_0)^3}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2=\frac{r^2}{8(n_1-n_0)^4}\left[\frac{n_1}{s'}-\frac{n_0}{s}\right]\left[(p_0-p_1)^2+(q_0-q_1)^2\right]^2.$$

Подставляя это выражение в (2), получим

$$T^{(4)} = \frac{1}{8n_0 s}\left\{\frac{n_0 r}{(n_1-n_0)^2}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]-\frac{s}{n_0}(p_0^2+q_0^2)\right\}^2 - \\ -\frac{1}{8n_1 s'}\left\{\frac{n_1 r}{(n_1-n_0)^2}\left[(p_0-p_1)^2+(q_0-q_1)^2\right]-\frac{s'}{n_1}(p_1^2+q_1^2)\right\}^2 - \\ -\frac{br}{8(n_1-n_0)^3}\left\{(p_0-p_1)^2+(q_0-q_1)^2\right\}^2. \tag{6}$$

Переменные в (6) можно заменить на их параксиальные значения; в частности, переменные Зайделя, относящиеся к точкам на падающем и преломленном лучах, можно поменять местами. Тогда для получения $\psi^{(4)}$ как функции

*) Как и в § 4.1, r обозначает радиус кривизны преломляющей поверхности. Его не следует путать с символом $\mathbf{r}$, который обозначает инвариант вращения $\sqrt{x_0^2+y_0^2}$, введенный в (5.3.2).

x_0, y_0, ξ_1 и η_1 можно использовать вместо (5.2.9) соотношения

$$\left.\begin{aligned} p_0 &= \frac{n_0\lambda_0}{D_0}\xi_1 - \frac{1}{\lambda_0}x_0, & p_1 &= \frac{n_1\lambda_1}{D_1}\xi_1 - \frac{1}{\lambda_1}x_0, \\ q_0 &= \frac{n_0\lambda_0}{D_0}\eta_1 - \frac{1}{\lambda_0}y_0, & q_1 &= \frac{n_1\lambda_1}{D_1}\eta_1 - \frac{1}{\lambda_1}y_0. \end{aligned}\right\} \tag{7}$$

Полезно провести еще одну модификацию. Гауссовы поперечные увеличения предмет — изображение (l_1/l_0) и входной — выходной зрачки (λ_1/λ_0) можно определить из (4.4.14) и (4.4.10) или просто из тех соображений, что отображение сферической поверхностью является проекцией из центра соответствующей сферы. Тогда с учетом (4) получим

$$\frac{l_1}{l_0} = \frac{r-s'}{r-s} = \frac{n_0 s'}{n_1 s}, \qquad \frac{\lambda_1}{\lambda_0} = \frac{r-t'}{r-t} = \frac{n_0 t'}{n_1 t}. \tag{8}$$

Вводя обозначения

$$h = \frac{\lambda_0 s}{D_0} = \frac{\lambda_1 s'}{D_1}, \qquad H = \frac{t}{\lambda_0 n_0} = \frac{t'}{\lambda_1 n_1}, \tag{9}$$

где были использованы соотношения (8) и (5.2.7), перепишем (7) в виде

$$\left.\begin{aligned} p_0 &= n_0\left(\frac{h}{s}\xi_1 - \frac{H}{t}x_0\right), & p_1 &= n_1\left(\frac{h}{s'}\xi_1 - \frac{H}{t'}x_0\right), \\ q_0 &= n_0\left(\frac{h}{s}\eta_1 - \frac{H}{t}y_0\right), & q_1 &= n_1\left(\frac{h}{s'}\eta_1 - \frac{H}{t'}y_0\right). \end{aligned}\right\} \tag{10}$$

Если, как и раньше, обозначить через r^2, ρ^2 и κ^2 три инварианта вращения

$$\mathrm{r}^2 = x_0^2 + y_0^2, \quad \rho^2 = \xi_1^2 + \eta_1^2, \quad \kappa^2 = x_0\xi_1 + y_0\eta_1, \tag{11}$$

то члены в фигурных скобках в соотношении (6) примут вид

$$(p_0 - p_1)^2 + (q_0 - q_1)^2 = \left(\frac{n_1 - n_0}{r}\right)^2 [H^2\mathrm{r}^2 + h^2\rho^2 - 2hH\kappa^2],$$

$$\frac{n_0 r}{(n_1-n_0)^2}\left[(p_0-p_1)^2 + (q_0-q_1)^2\right] - \frac{s}{n_0}(p_0^2 + q_0^2) = \\ = H^2\mathrm{r}^2\left[L - (K-L)\frac{s}{t}\right] + h^2\rho^2 K - 2hH\kappa^2 L,$$

$$\frac{n_1 r}{(n_1-n_0)^2}\left[(p_0-p_1)^2 + (q_0-q_1)^2\right] - \frac{s'}{n_1}(p_1^2 + q_1^2) = \\ = H^2\mathrm{r}^2\left[L - (K-L)\frac{s'}{t'}\right] + h^2\rho^2 K - 2hH\kappa^2 L.$$

Подставляя эти соотношения в (6) и используя (1), получим окончательное выражение для $\psi^{(4)}$

$$\psi^{(4)} = \frac{1}{8}\mathrm{r}^4 H^4\left\{\frac{b}{r^3}(n_0 - n_1) + L^2\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right) - \right. \\ \left. - 2L(K-L)\left(\frac{1}{n_0 t} - \frac{1}{n_1 t'}\right) + (K-L)^2\left(\frac{s}{n_0 t^2} - \frac{s'}{n_1 t'^2}\right)\right\} + \\ + \frac{1}{8}\rho^4 h^4\left\{\frac{b}{r^3}(n_0 - n_1) + K^2\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right)\right\} + \\ + \frac{1}{2}\kappa^4 H^2 h^2\left\{\frac{b}{r^3}(n_0 - n_1) + L^2\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right)\right\} + \\ + \frac{1}{2}\mathrm{r}^2\rho^2 H^2 h^2\left\{\frac{b}{r^3}(n_0 - n_1) + KL\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right) - \right. \\ \left. - K(K-L)\left(\frac{1}{n_0 t} - \frac{1}{n_1 t'}\right)\right\} - \frac{1}{2}\mathrm{r}^2\kappa^2 H^3 h\left\{\frac{b}{r^3}(n_0 - n_1) + L^2\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right) - \right. \\ \left. - L(K-L)\left(\frac{1}{n_0 t} - \frac{1}{n_1 t'}\right)\right\} - \frac{1}{2}\rho^2\kappa^2 H h^3\left\{\frac{b}{r^3}(n_0 - n_1) + KL\left(\frac{1}{n_0 s} - \frac{1}{n_1 s'}\right)\right\}. \tag{12}$$

Эта формула, если ее сравнить с общим выражением (5.3.3), дает коэффициенты разложения четвертого порядка A, B, ..., F возмущенного эйконала преломляющей поверхности вращения.

Теперь нетрудно обобщить этот результат на случай произвольной центрированной системы, состоящей из любого числа преломляющих поверхностей. Отметим индексом i величины, относящиеся к i-й поверхности, и пусть n_i — показатель преломления среды, расположенной за i-й поверхностью. Тогда, используя теорему сложения из § 5.4 и сравнивая (12) с формулой (5.3.3), находим

$$\left.\begin{aligned}
B &= \frac{1}{2}\sum_i h_i^4 \left\{\frac{b_i}{r_i^3}(n_i - n_{i-1}) + K_i^2\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1}s_i}\right)\right\},\\
C &= \frac{1}{2}\sum_i H_i^2 h_i^2 \left\{\frac{b_i}{r_i^3}(n_i - n_{i-1}) + L_i^2\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1}s_i}\right)\right\},\\
D &= \frac{1}{2}\sum_i H_i^2 h_i^2 \left\{\frac{b_i}{r_i^3}(n_i - n_{i-1}) + K_i L_i\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1}s_i}\right) - \right.\\
&\qquad\left. - K_i(K_i - L_i)\left(\frac{1}{n_i t_i'} - \frac{1}{n_{i-1}t_i}\right)\right\},\\
E &= \frac{1}{2}\sum_i H_i^3 h_i \left\{\frac{b_i}{r_i^3}(n_i - n_{i-1}) + L_i^2\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1}s_i}\right) - \right.\\
&\qquad\left. - L_i(K_i - L_i)\left(\frac{1}{n_i t_i'} - \frac{1}{n_{i-1}t_i}\right)\right\},\\
F &= \frac{1}{2}\sum_i H_i h_i^3 \left\{\frac{b_i}{r_i^3}(n_i - n_{i-1}) + K_i L_i\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1}s_i}\right)\right\}.
\end{aligned}\right\} \quad (13)$$

Это и есть *формулы Зайделя* для коэффициентов первичных аберраций произвольной центрированной системы преломляющих поверхностей *).

Соотношения (13) выражают коэффициенты первичных аберраций через величины, характеризующие прохождение двух параксиальных лучей через систему, а именно одного луча, выходящего из осевой точки предмета, и другого — выходящего из центра входного зрачка. Полезно сделать сводку соответствующих формул Гаусса. Пусть d_i — расстояние между полюсами i-й и $(i+1)$-й поверхностей. Поскольку параксиальное изображение первыми i поверхностями системы служит предметом для $(i+1)$-й поверхности, можно написать следующие формулы перехода:

$$s_{i+1} = s_i' - d_i, \qquad t_{i+1} = t_i' - d_i. \quad (14)$$

При заданных расстояниях s_1 и t_1 от плоскости предмета и плоскости входного зрачка до полюса первой поверхности можно с помощью соотношений Аббе

$$\left.\begin{aligned}
n_{i-1}(1/r_i - 1/s_i) &= n_i(1/r_i - 1/s_i') = K_i,\\
n_{i-1}(1/r_i - 1/t_i) &= n_i(1/r_i - 1/t_i') = L_i
\end{aligned}\right\} \quad (15)$$

и формулы (14) последовательно вычислить расстояние s_1', t_1', s_2, t_2, s_2', t_2', ... и соответствующие значения K_1, L_1, K_2, L_2, ... Необходимо также определить величины h_i и H_i. Если положить для простоты длину λ_0 в плоскости входного зрачка равной *единице* и использовать соотношения $D_i = t_i' - s_i' = t_{i+1} - s_{i+1}$, то из (9) следует, что величины h_i и H_i можно последовательно вычислить с

*) Фактически Зайдель исследовал системы, состоящие только из сферических поверхностей. Эффекты несферичности (характеризующиеся постоянными b_i) были учтены позднее (см. [17]).

помощью уравнений *)

$$H_1=\frac{t_1}{n_0},\qquad h_1=\frac{s_1}{t_1-s_1},\qquad \frac{H_{i+1}}{H_i}=\frac{t_{i+1}}{t_i'},\qquad \frac{h_{i+1}}{h_i}=\frac{s_{i+1}}{s_i'}. \tag{16}$$

Из формул (9) и из соотношений Аббе (4) и (5) находим следующее выражение, которое может служить для проверки вычислений и понадобится нам позже:

$$h_iH_i=\frac{s_it_i}{n_{i-1}(t_i-s_i)}=\frac{s_i't_i'}{n_i(t_i'-s_i')}=\frac{1}{L_i-K_i}. \tag{17}$$

5.5.2. Формулы Зайделя, выраженные через параметры одного параксиального луча. Часто представляется желательным выразить коэффициенты первичной аберрации в такой форме, которая наиболее отчетливо показывает их зависимость от параметров оптической системы; формулы (13) непригодны для этой цели. Чтобы воспользоваться ими необходимо, в соответствии с законами параксиальной оптики, построить ход двух лучей, один из которых выходит из осевой точки предмета, а другой — из центра входного зрачка. Зайдель, однако, показал, что параметры второго луча можно исключить, т. е. выразить коэффициенты первичных аберраций только через параметры первого луча. Естественно, в формулах останется величина, зависящая от положения входного зрачка, поскольку ясно, что изменение положения соответствующей диафрагмы влияет на аберрации.

Для исключения величин, относящихся к лучу из центра входного зрачка, необходимо прежде всего выразить расстояния от оси H_i через h_i. Из (14), (16) и (17) имеем

$$\frac{H_{i+1}}{h_{i+1}}-\frac{H_i}{h_i}=\frac{1}{h_ih_{i+1}}H_ih_i\left(\frac{t_{i+1}}{t_i'}-\frac{s_{i+1}}{s_i'}\right)=\frac{1}{h_ih_{i+1}}H_ih_id_i\left(\frac{1}{s_i'}-\frac{1}{t_i'}\right)=\frac{d_i}{n_ih_ih_{i+1}}. \tag{18}$$

Если положить

$$H_i=k_ih_i, \tag{19}$$

то из (18) следует, что

$$k_{i+1}=k_1+\sum_{j=1}^{i}\frac{d_j}{n_jh_jh_{j+1}}, \tag{20}$$

где, согласно (16),

$$k_1=\frac{H_1}{h_1}=\frac{t_1(t_1-s_1)}{n_0s_1}. \tag{21}$$

Другие величины в (13), характеризующие параксиальный луч из центра входного зрачка, также легко выразить через параметры другого луча. Из (17) и (19) имеем

$$L_i=K_i+\frac{1}{k_ih_i^2}. \tag{22}$$

Далее в соответствии с (15) и (22) мы можем написать

$$\frac{1}{n_it_i'}-\frac{1}{n_{i-1}t_i}=\frac{1}{n_ir_i}-\frac{1}{n_{i-1}r_{i-1}}-\left(\frac{1}{n_i^2}-\frac{1}{n_{i-1}^2}\right)L_i=$$
$$=\frac{1}{n_i}\left(\frac{K_i}{n_i}+\frac{1}{s_i'}\right)-\frac{1}{n_{i-1}}\left(\frac{K_i}{n_{i-1}}+\frac{1}{s_i}\right)-\left(\frac{1}{n_i^2}-\frac{1}{n_{i-1}^2}\right)\left(K_i+\frac{1}{k_ih_i^2}\right)=$$
$$=\frac{1}{n_is_i'}-\frac{1}{n_{i-1}s_i}-\frac{1}{k_ih_i^2}\left(\frac{1}{n_i^2}-\frac{1}{n_{i-1}^2}\right). \tag{23}$$

---

*) Из соотношения $h_{i+1}/h_i=s_{i+1}/s_i'$ следует, что h_i пропорционально расстоянию от оси, на котором параксиальный луч, выходящий из осевой точки предмета, пересекает i-ю поверхность. Соотношение $H_{i+1}/H_i=t_{i+1}/t_i'$ интерпретируется таким же образом для параксиального луча, выходящего из центра входного зрачка.

Подставляя (19), (22) и (23) в (13), окончательно получим

$$\left.\begin{aligned}
B &= \frac{1}{2}\sum_i h_i^4 \frac{b_i}{r_i^3}(n_i - n_{i-1}) + h_i^4 K_i^2 \left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right),\\
C &= \frac{1}{2}\sum_i h_i^4 k_i^2 \frac{b_i}{r_i^3}(n_i - n_{i-1}) + (1 + h_i^2 k_i K_i)^2 \left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right),\\
D &= \frac{1}{2}\sum_i h_i^4 k_i^2 \frac{b_i}{r_i^3}(n_i - n_{i-1}) + \\
&\quad + h_i^2 k_i K_i (2 + h_i^2 k_i K_i)\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right) - K_i\left(\frac{1}{n_i^2} - \frac{1}{n_{i-1}^2}\right),\\
E &= \frac{1}{2}\sum_i h_i^4 k_i^3 \frac{b_i}{r_i^3}(n_i - n_{i-1}) + k_i (1 + h_i^2 k_i K_i)(2 + h_i^2 k_i K_i) \times \\
&\quad \times \left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right) - \frac{1 + h_i^2 k_i K_i}{h_i^2}\left(\frac{1}{n_i^2} - \frac{1}{n_{i-1}^2}\right),\\
F &= \frac{1}{2}\sum_i h_i^4 k_i \frac{b_i}{r_i^3}(n_i - n_{i-1}) + h_i^2 K_i (1 + k_i h_i^2 K_i)\left(\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right).
\end{aligned}\right\} \quad (24)$$

Это и есть требуемая форма соотношений Зайделя. Координаты входного зрачка содержатся только в коэффициенте k_1, который связан с расстоянием l_1 от плоскости входного зрачка до полюса первой поверхности выражением (21). Величины s_i, s_i', K_i и h_i вычисляются, как и раньше, с помощью соответствующих формул (14) — (16), а величины k_i — с помощью (20).

5.5.3. Теорема Петцваля. Из выражений для коэффициентов астигматизма и кривизны поля можно вывести интересное соотношение, полученное впервые Петцвалем. Учитывая (15), имеем

$$\begin{aligned}
C - D = \frac{1}{2}\sum_i \Big\{&[(1 + h_i^2 k_i K_i)^2 - 2h_i^2 k_i K_i - h_i^4 k_i^2 K_i^2]\left[\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i}\right] + \\
&+ K_i\left(\frac{1}{n_i^2} - \frac{1}{n_{i-1}^2}\right)\Big\} = \frac{1}{2}\sum_i \left\{\frac{1}{n_i s_i'} - \frac{1}{n_{i-1} s_i} + K_i\left(\frac{1}{n_i^2} - \frac{1}{n_{i-1}^2}\right)\right\} = \\
&= \frac{1}{2}\sum_i \frac{1}{r_i}\left(\frac{1}{n_i} - \frac{1}{n_{i-1}}\right). \qquad (25)
\end{aligned}$$

Согласно § 5.3 коэффициенты C и D определяют сагиттальную и тангенциальную кривизны поля. Если обозначить через n_a показатель преломления последней среды, то из уравнений (5.3.18) и (5.3.19) получим

$$C = \frac{1}{4n_a}\left(\frac{1}{R_t} - \frac{1}{R_s}\right), \quad D = \frac{1}{2n_a}\frac{1}{R_s}. \qquad (26)$$

Следовательно, (25) можно представить в виде

$$\frac{1}{R_t} - \frac{3}{R_s} = 2n_a \sum \frac{1}{r_i}\left(\frac{1}{n_i} - \frac{1}{n_{i-1}}\right). \qquad (27)$$

Таким образом, мы получили соотношение, связывающее кривизны двух фокальных поверхностей, которое содержит только радиусы кривизны преломляющих поверхностей системы и соответствующие показатели преломления. Если система свободна от сферической аберрации, комы и астигматизма, то на поверхности радиуса $R_s = R_t = R$ получается резкое изображение; радиус этой поверхности, согласно (27), можно найти из соотношения

$$\frac{1}{R} = -n_a \sum_i \frac{1}{r_i}\left(\frac{1}{n_i} - \frac{1}{n_{i-1}}\right). \qquad (28)$$

Полученный результат называется *теоремой Петцваля*.

Условие

$$\sum_i \frac{1}{r_i}\left(\frac{1}{n_i}-\frac{1}{n_{i-1}}\right)=0 \tag{29}$$

называется *условием Петцваля.* Оно служит необходимым условием того, что *поле будет плоским.* Однако следует помнить, что оно справедливо лишь в рамках теории Зайделя; вне этих рамок оно теряет свое значение.

Сферическая поверхность, касающаяся двух фокальных поверхностей в их общей осевой точке и имеющая радиус R, определенный (28), называется *поверхностью Петцваля* независимо от наличия аберраций. Согласно (27) и (28) радиусы кривизны сагиттальной фокальной поверхности, тангенциальной фокальной поверхности и поверхности Петцваля связаны между собой соотношением

$$\frac{3}{R_s}-\frac{1}{R_t}=\frac{2}{R}. \tag{30}$$

§ 5.6. Пример: первичные аберрации тонкой линзы

Используем теперь формулы Зайделя для нахождения коэффициентов первичных аберраций тонкой линзы с показателем преломления n, расположенной в воздухе (вакууме). В этом случае

$$n_0=n_2=1, \qquad n_1=n. \tag{1}$$

Поскольку толщиной линзы d можно пренебречь, то, согласно (5.5.14), имеем

$$s_2=s_1', \tag{2}$$

и (5.5.15) дает

$$\left.\begin{aligned} 1/r_1-1/s_1&=n\,(1/r_1-1/s_1')=K_1,\\ n\,(1/r_2-1/s_2)&=(1/r_2-1/s_2')=K_2. \end{aligned}\right\} \tag{3}$$

Далее, учитывая (2), получим из (5.5.16) и (5.5.20)

$$\left.\begin{aligned} &h_1=h_2=h, \qquad k_1=k_2=k,\\ &h=\frac{s_1}{t_1-s_1}, \qquad k=\frac{t_1(t_1-s_1)}{s_1}. \end{aligned}\right\} \tag{4}$$

где

Линзу удобно характеризовать несколькими простыми параметрами. Пусть $\mathcal{P}$ — оптическая сила линзы (см. уравнения (4.4.35) и (4.4.36)), т. е.

$$\mathcal{P}=1/f=(n-1)(1/r_1-1/r_2), \tag{5}$$

и пусть

$$\sigma=(n-1)(1/r_1+1/r_2). \tag{6}$$

Из (3) имеем

$$(n-1)\frac{1}{r_1}=\frac{n}{s_1'}-\frac{1}{s_1}, \qquad (n-1)\frac{1}{r_2}=\frac{n}{s_2}-\frac{1}{s_2'}, \tag{7}$$

так что с учетом (2)

$$1/s_2'-1/s_1=\mathcal{P}. \tag{8}$$

Перепишем это соотношение в виде

$$-1/s_1-\mathcal{P}/2=-1/s_2'+\mathcal{P}/2=\mathcal{K}; \tag{9}$$

здесь $\mathcal{K}$ играет для линзы ту же роль, что и K для одной поверхности. Соответственно величину $\mathcal{K}$ мы будем называть *инвариантом Аббе линзы.*

Позднее будет показано, что коэффициенты деформации b_1 и b_2 двух поверхностей линзы входят в формулы лишь в комбинации

$$\beta=(n-1)\left(\frac{b_1}{r_1^3}-\frac{b_2}{r_2^3}\right); \tag{10}$$

эту величину можно назвать *коэффициентом деформации линзы.*

Выразим теперь величины, входящие в формулы Зайделя (5.5.24), через параметры $\mathcal{P}$, σ, β и $\mathcal{K}$. Первые три параметра характеризуют линзу, а последний — положение предмета. Сначала получим из (3)

$$K_1=\mathcal{K}+\frac{\sigma+n\mathcal{P}}{2(n-1)},\qquad K_2=\mathcal{K}+\frac{\sigma-n\mathcal{P}}{2(n-1)}. \tag{11}$$

Далее

$$\left.\begin{aligned}\frac{1}{n_1s_1'}-\frac{1}{s_1}&=\frac{n^2-1}{n^2}\mathcal{K}+\frac{\sigma}{2n^2}+\frac{\mathcal{P}}{2},\\ \frac{1}{s_2'}-\frac{1}{ns_2}&=-\frac{n^2-1}{n^2}\mathcal{K}-\frac{\sigma}{2n^2}+\frac{\mathcal{P}}{2}.\end{aligned}\right\} \tag{12}$$

Подставляя эти соотношения в формулы Зайделя (5.5.24), находим следующие выражения для коэффициентов первичных аберраций тонкой линзы:

$$\left.\begin{aligned}B&=h^4U,\\ F&=h^4kU+h^2V,\\ C&=h^4k^2U+2h^2k^2V+\frac{1}{2}\mathcal{P},\\ D&=h^4k^2U+2h^2k^2V+\frac{n+1}{2n}\mathcal{P},\\ E&=h^4k^3U+3h^2k^2V+k\frac{3n+1}{2n}\mathcal{P},\end{aligned}\right\} \tag{13}$$

где

$$\left.\begin{aligned}U&=\frac{1}{2}\beta+\frac{n^2}{8(n-1)^2}\mathcal{P}^3-\frac{n}{2(n+2)}\mathcal{K}^2\mathcal{P}+\\ &\qquad+\frac{1}{2n(n+2)}\mathcal{P}\left[\frac{n+2}{2(n-1)}\sigma+2(n+1)\mathcal{K}\right]^2,\\ V&=\frac{1}{2n}\mathcal{P}\left[\frac{n+1}{2(n-1)}\sigma+(2n+1)\mathcal{K}\right].\end{aligned}\right\} \tag{14}$$

Здесь мы рассмотрим только случай совпадения *входного зрачка с линзой* ($t_1=0$). Согласно (4) имеем

$$h=-1,\quad k=0,$$

и формулы (13) принимают вид

$$\left.\begin{aligned}&B=U, && F=V,\\ &\frac{1}{2}(C+D)=\frac{2n+1}{4n}\mathcal{P}, && \frac{1}{2}(C-D)=-\frac{1}{4n}\mathcal{P},\\ &E=0.\end{aligned}\right\} \tag{15}$$

Таким образом, величины U и V, определенные (14), характеризуют сферическую аберрацию и кому линзы с диафрагмой, расположенной в ее плоскости. В такой линзе отсутствует дисторсия ($E=0$), однако всегда имеются астигматизм и кривизна поля ($C\neq 0$, $D\neq 0$).

Выясним, существует ли для такой линзы пара апланатических точек ($B=F=0$). Согласно (15) кома будет отсутствовать, если $V=0$, или

$$\frac{n-1}{2(n-1)}\sigma+(2n+1)\mathcal{K}=0. \tag{16}$$

Тогда из (9) найдем расстояние до предмета s_1

$$\frac{1}{s_1}=-\frac{1}{2}\mathcal{P}+\frac{n+1}{2(n-1)(2n+1)}\sigma. \tag{17}$$

При таком выборе s_1 сферическая аберрация линзы полностью определена. Следовательно, в общем случае для линзы с диафрагмой, лежащей в плоскости линзы, не существует пары апланатических точек. Однако если заданы только $\mathcal{P}$ и σ, то сферическую аберрацию можно устранить, выбрав подходящий коэффициент деформации β.

Если же заданы s_1 и фокусное расстояние $f = 1/\mathcal{P}$, то можно так подобрать радиусы r_1 и r_2, чтобы исчезла кома. В этом случае, согласно (5), (6) и (17), имеем

$$\left.\begin{aligned}\frac{1}{r_1}&=\frac{2n+1}{n+1}\frac{1}{s_1}+\frac{n^2}{n^2-1}\mathcal{P},\\ \frac{1}{r_2}&=\frac{2n+1}{n+1}\frac{1}{s_1}+\frac{n^2-n-1}{n^2-1}\mathcal{P}.\end{aligned}\right\} \tag{18}$$

Если, например, предмет находится в бесконечности ($s_1=\infty$) и $n = 1{,}5$, то

$$r_1=\frac{5}{9}\frac{1}{\mathcal{P}}=\frac{5}{9}f, \qquad r_2=-\frac{5}{\mathcal{P}}=-5f. \tag{19}$$

На рис. 5.10 изображена такая линза, свободная от комы.

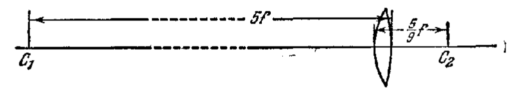


Рис. 5.10. Тонкая линза, свободная от комы.
Предмет находится в бесконечности, C_1 и C_2 — центры кривизны поверхностей линзы.

Сферическую аберрацию можно теперь устранить подходящим выбором β. Если предмет находится в бесконечности, то, согласно (9), $\mathcal{K} = -\mathcal{P}/2$ и, следовательно, из (16) имеем $\sigma = (2n+1)(n-1)\mathcal{P}/(n+1)$; тогда из условия отсутствия сферической аберрации $U = 0$ получим следующее выражение для β:

$$-\frac{1}{2}\beta=\mathcal{P}^3\left\{\frac{n^2}{8(n-1)^2}-\frac{n}{8(n+2)}+\frac{1}{2n(n+2)}\left[\frac{(n+2)(2n+1)}{2(n+1)}-(n+1)^2\right]\right\}, \tag{20}$$

или

$$\beta=-\frac{n^3}{(n^2-1)^2}\mathcal{P}^3.$$

При $n = 1{,}5$ имеем

$$\beta=-\frac{54}{25}\mathcal{P}^3, \tag{21}$$

так что из (10) и (19) находим

$$729b_1+b_2=-540. \tag{22}$$

Последнее соотношение выполняется, например, при $b_1 = b_2 = -0{,}74$.

Можно вместо устранения комы попытаться сначала уменьшить сферическую аберрацию. Оказывается, что в случае положительного фокусного расстояния линзы сферическая аберрация исчезает полностью лишь тогда, когда предмет находится в определенной, ограниченной области пространства; при $n = 1{,}5$ эта область простирается от $0{,}36\,f$ до $0{,}44\,f$ за линзой. При любом другом положении предмета некоторая сферическая аберрация всегда имеется.

Для устранения кривизны поля необходимо использовать диафрагму, помещенную в соответствующее место, которое можно определить из общих формул (13). Для линзы с положительным фокусным расстоянием и $n = 1{,}5$ удается устранить кривизну поля, только если предмет находится в области, простирающейся от точки, удаленной на одно фокусное расстояние перед линзой, до точки, удаленной на половину фокусного расстояния за линзой.

Для системы, состоящей из нескольких линз, вычисления становятся значительно сложнее и основная задача состоит в устранении хроматической аберрации. Относительно просто исследовать первичные аберрации ахроматического объектива телескопа, состоящего из двух склеенных тонких линз; оказывается, что в этом случае все первичные аберрации, кроме астигматизма и и кривизны поля, можно устранить, следовательно, такой объектив годится только для узкого поля зрения (не более 3° для системы $f/10$).

§ 5.7. Хроматическая аберрация произвольной центрированной системы линз

Хроматическая аберрация была уже кратко рассмотрена в § 4.7. Здесь же с помощью общего формализма, развитого в предыдущих параграфах *), будут получены в явном виде выражения для хроматизма положения изображения и хроматизма увеличения системы. Предположим, что для монохроматического света система полностью скорректирована. Такое предположение допустимо, если исследуются только основные эффекты, так как можно считать, что изменения монохроматических аберраций при небольших изменениях длины волны малы по сравнению с самими аберрациями; в общем случае эти изменения имеют те же порядки величин, что и члены, отброшенные в теории Зайделя.

Согласно (5.5.14) и (5.5.15) имеем

$$n_{i-1}(1/r_i - 1/s_i) = n_i(1/r_i - 1/s_i') = K_i, \tag{1}$$

$$s_{i+1} = s_i' - d_i. \tag{2}$$

Если обозначать изменение какой-либо величины при малом изменении $\delta\lambda$ длины волны через δ, то из (1) и (2) находим

$$n_i \frac{\delta s_i'}{s_i'^2} - n_{i-1}\frac{\delta s_i}{s_i^2} = -\left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right)K_i, \tag{3}$$

$$\delta s_{i+1} = \delta s_i'. \tag{4}$$

Умножая (3) на h_i^2 и используя соотношение

$$\frac{h_{i+1}}{h_i} = \frac{s_{i+1}}{s_i'}, \tag{5}$$

получим

$$n_i\left(\frac{h_{i+1}}{s_{i+1}}\right)^2 \delta s_{i+1} - n_{i-1}\left(\frac{h_i}{s_i}\right)^2 \delta s_i = -h_i^2 K_i\left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right). \tag{6}$$

Пусть α — число поверхностей в исследуемой оптической системе. Тогда, складывая все уравнения вида (6), имеем

$$n_\alpha\left(\frac{h_{\alpha+1}}{s_{\alpha+1}}\right)^2 \delta s_{\alpha+1} = n_\alpha\left(\frac{h_\alpha}{s_\alpha'}\right)^2 \delta s_\alpha' = n_0\left(\frac{h_1}{s_1}\right)^2 \delta s_1 - \sum_{i=1}^{\alpha} h_i^2 K_i\left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right).$$

Так как положение предмета не зависит от длины волны, $\delta s_1 = 0$, и мы получаем следующее выражение для *хроматической аберрации* δs_α *положения изображения*:

$$\delta s_\alpha' = -\frac{1}{n_\alpha}\left(\frac{s_\alpha'}{h_\alpha}\right)^2 \sum_{i=1}^{\alpha} h_i^2 K_i\left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right). \tag{7}$$

Рассмотрим теперь хроматическую аберрацию увеличения системы M. Имеем

$$M = \frac{l_1}{l_0}\cdot\frac{l_2}{l_1}\cdot\frac{l_3}{l_2}\cdots\frac{l_\alpha}{l_{\alpha-1}}. \tag{8}$$

Согласно (5.5.8)

$$\frac{l_i}{l_{i-1}} = \frac{n_{i-1}}{n_i}\frac{s_i'}{s_i},$$

и поэтому

$$M = \frac{n_0}{n_\alpha}\left(\frac{s_1'}{s_1}\frac{s_2'}{s_2}\frac{s_3'}{s_3}\cdots\frac{s_\alpha'}{s_\alpha}\right). \tag{9}$$

*) Здесь мы снова, по существу, следуем анализу Шварцшильда [6].

Беря логарифмическую производную, получим

$$\frac{\delta M}{M}=\frac{\delta n_0}{n_0}-\frac{\delta n_\alpha}{n_\alpha}+\sum_{i=1}^{\alpha}\left(\frac{\delta s_i'}{s_i'}-\frac{\delta s_i}{s_i}\right)=$$

$$=\left(\frac{\delta n_0}{n_0}-\frac{\delta n_\alpha}{n_\alpha}\right)+\frac{\delta s_\alpha'}{s_\alpha'}-\frac{\delta s_1}{s_1}-\sum_{i=2}^{\alpha}\delta s_i\left(\frac{1}{s_i}-\frac{1}{s_{i-1}'}\right). \quad (10)$$

Сумму, входящую в это выражение, перепишем в виде

$$\sum_{i=2}^{\alpha}\delta s_i\left(\frac{1}{s_i}-\frac{1}{s_{i-1}'}\right)=\sum_{i=2}^{\alpha}\frac{d_{i-1}\delta s_i}{s_i s_{i-1}'}=\sum_{i=2}^{\alpha}\frac{d_{i-1}}{s_i^2}\frac{h_i}{h_{i-1}}\delta s_i.$$

Но из (5.5.18) следует, что

$$\frac{d_{i-1}}{h_{i-1}}=(k_i-k_{i-1})\,h_i n_{i-1}$$

и, следовательно,

$$\sum_{i=2}^{\alpha}\delta s_i\left(\frac{1}{s_i}-\frac{1}{s_{i-1}'}\right)=\sum_{i=2}^{\alpha}h_i^2 n_{i-1}\left(\frac{k_i-k_{i-1}}{s_i^2}\right)\delta s_i,$$

или, перегруппировав члены с k_i,

$$\sum_{i=2}^{\alpha}\delta s_i\left(\frac{1}{s_i}-\frac{1}{s_{i-1}'}\right)=-n_0\left(\frac{h_1}{s_1}\right)^2 k_1\delta s_1+n_\alpha\left(\frac{h_{\alpha+1}}{s_{\alpha+1}}\right)^2 k_\alpha\delta s_{\alpha+1}-$$

$$-\sum_{i=1}^{\alpha}k_i\left[n_i\left(\frac{h_{i+1}}{s_{i+1}}\right)^2\delta s_{i+1}-n_{i-1}\left(\frac{h_i}{s_i}\right)^2\delta s_i\right].$$

Переписав второй член в правой части этого соотношения с помощью (4) и (5), а последний — с помощью (6), получим

$$\sum_{i=2}^{\alpha}\delta s_i\left(\frac{1}{s_i}-\frac{1}{s_{i-1}'}\right)=-n_0\left(\frac{h_1}{s_1}\right)^2 k_1\delta s_1+n_\alpha\left(\frac{h_\alpha}{s_\alpha'}\right)^2 k_\alpha\delta s_\alpha'+$$

$$+\sum_{i=1}^{\alpha}h_i^2 k_i K_i\left(\frac{\delta n_i}{n_i}-\frac{\delta n_{i-1}}{n_{i-1}}\right).$$

Подставляя это выражение в (10) и используя тот факт, что $\delta s_1=0$, имеем

$$\frac{\delta M}{M}=\left(\frac{\delta n_0}{n_0}-\frac{\delta n_\alpha}{n_\alpha}\right)+\frac{\delta s_\alpha'}{s_\alpha'}\left(1-\frac{n_\alpha}{s_\alpha'}h_\alpha^2 k_\alpha\right)-\sum_{i=1}^{\alpha}h_i^2 k_i K_i\left(\frac{\delta n_i}{n_i}-\frac{\delta n_{i-1}}{n_{i-1}}\right).$$

Однако, согласно (5.5.15) и (5.5.22)

$$\frac{\delta s_\alpha'}{s_\alpha'}\left[1-\frac{n_\alpha}{s_\alpha'}h_\alpha^2 k_\alpha\right]=\frac{\delta s_\alpha'}{s_\alpha'}\left[1-\frac{n_\alpha}{s_\alpha'(L_\alpha-K_\alpha)}\right]=\frac{\delta s_\alpha'}{s_\alpha'-t_\alpha'},$$

и, следовательно, окончательное выражение для *хроматического изменения* δM *увеличения системы* будет иметь вид

$$\frac{\delta M}{M}=\left(\frac{\delta n_0}{n_0}-\frac{\delta n_\alpha}{n_\alpha}\right)+\frac{\delta s_\alpha'}{s_\alpha'-t_\alpha'}-\sum_{i=1}^{\alpha}h_i^2 k_i K_i\left(\frac{\delta n_i}{n_i}-\frac{\delta n_{i-1}}{n_{i-1}}\right). \quad (11)$$

Если, как это обычно бывает на практике, показатели преломления сред в пространствах предмета и изображения одинаковы, то первые два члена в правой части (11) пропадают. Если система ахроматизирована относительно положения изображения, то $\delta s_\alpha'$ (определенная в (7)) тоже исчезает, и мы получаем следующие два условия, обеспечивающие отсутствие хроматических искажений

положения изображения и увеличения системы:

$$\sum_{i=1}^{\alpha} h_i^2 K_i \left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right) = 0 \tag{12}$$

и

$$\sum_{i=1}^{\alpha} h_i^2 k_i K_i \left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right) = 0. \tag{13}$$

Для тонкой линзы, находящейся в воздухе, мы можем, используя соответствующие формулы из § 5.6, написать эти суммы в виде

$$\sum h_i^2 K_i \left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right) = h^2 \mathcal{P} \frac{\delta n}{n-1} \tag{14}$$

и

$$\sum h_i^2 k_i K_i \left(\frac{\delta n_i}{n_i} - \frac{\delta n_{i-1}}{n_{i-1}}\right) = h^2 k \mathcal{P} \frac{\delta n}{n-1}, \tag{15}$$

где h, k и оптическая сила $\mathcal{P}$ линзы определены (4), (5.6.4) и (5.6.5). Тогда, подставляя (14) в (7) и (15) в (11), найдем соответствующие выражения для хроматических искажений положения изображения и увеличения линзы.

ЛИТЕРАТУРА

1. J. Petzval, Bericht über die Ergebnisse einiger dioptrischer Untersuchungen, Pesth, 1843.
2. J. Petzval, Bericht über optische Untersuchungen, Ber. Kais. Akad. Wien, Math. naturwiss. Kl. 24, 50, 92, 129 (1857).
3. L. Seidel, Astr. Nachr. 43, № 1027, 289 (1856).
4. L. Seidel, Astr. Nachr. 43, № 1028, 305 (1856).
5. L. Seidel, Astr. Nachr. 43, № 1029, 321 (1856).
6. K. Schwarzschild, Abh. Königl. Ges. Wis. Göttingen, Math.-phys. Kl. 4, N 1, 2, 3 (1905—1906). Перепечатано в журнале Astr. Mitt. Königl. Sternwarte Göttingen (1905).
7. Г. Д. Рабинович, ЖЭТФ 16, 161 (1946).
8. E. Wolf, J. Opt. Soc. Amer. 42, 547 (1952).
9. J. L. Rayces, Optica Acta, 11, 85 (1964).
10. A. Kohlschütter, Dissertation, Göttingen, 1908.
11. M. Herzberger, J. Opt. Soc. Amer. 29, 395 (1939); Modern Geometrical Optics, Intersci. Publ. New York, 1958. (М. Герцбергер, Современная геометрическая оптика, ИЛ, 1962.)
12. J. Focke, в сб.: «Progr. in Optica», vol. 4, ed. E. Wolf, North Holland Publ. Co., Amsterdam; a. New. York, J. Wiley, a. Sons, N. Y., 1965, p. 1.
13. G. C. Steward, Trans. Cambr. Phil. Soc. 23, 235 (1926).
14. N. Chako, Trans. Chalmers Univ. Technology, Gothenburg № 191 (1957).
15. M. Rohr, Geometrical Investigation of Formation of Images in Optical Systems. Stationery Office, London, p. 344 (перевод с немецкого).

ГЛАВА 6

ОПТИЧЕСКИЕ ПРИБОРЫ, ФОРМИРУЮЩИЕ ИЗОБРАЖЕНИЕ

В трех предыдущих главах мы познакомились с геометрической теорией оптического отображения, пользуясь главным образом законами параксиальной оптики и теорией Зайделя. Исключительно велика ценность этого раздела оптики, позволяющего описывать принципы работы оптических приборов в наглядной форме. Качество оптических систем нельзя оценить при помощи одной только теории Гаусса, но она позволяет указать назначение отдельных оптических элементов, так что часто можно получить ясное, хотя и несколько упрощенное представление о действии системы без глубокого проникновения в сложную технику оптических расчетов.

Совершенствование оптических приборов в прошлом происходило непрерывно, по мере того, как преодолевались технические трудности. Однако два обстоятельства не позволяют дать последовательное описание конструкций оптических систем. Во-первых, простая теория не указывает на пределы применимости данного устройства; в каждом отдельном случае требуется длительный анализ, часто содержащий утомительные вычисления. Во-вторых, трудности практического характера часто могут препятствовать использованию полезного устройства. В приводимом ниже обзоре нет указаний на практические и теоретические положения, ограничивающие применение тех или иных устройств; в нем излагаются только основные принципы работы наиболее важных оптических приборов, что необходимо для понимания более строгой теории формирования оптического изображения, которой посвящены последующие главы *).

§ 6.1. Глаз

Вероятно, простейшим оптическим прибором можно считать устройство, состоящее из одной собирательной линзы, дающей действительное изображение предмета на светочувствительной поверхности. Примерами таких оптических систем служат фотоаппарат и глаз (рис. 6.1). Дно глазного яблока **), называемое *сетчаткой*, состоит из слоя светочувствительных элементов. Поскольку глаз часто входит как составная часть в оптическую систему, изучение его характеристик весьма существенно для инструментальной оптики. Этот параграф содержит описание некоторых особенностей глаза.

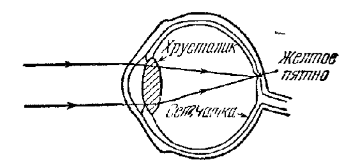


Рис. 6.1. Глаз человека.

Преломление света в глазе происходит главным образом на его внешней поверхности — *роговой оболочке*, или *роговице*, а также на поверхностях *хрусталика*, находящегося позади способной к сокращению *радужной оболочки* (*ириса*). Последняя определяет апертурный диаметр, величина которого в соответствии с полным световым потоком, попадающим в глаз, может изменяться непроизвольным мышечным усилием, приблизительно от 1,5 до 6 *мм*. Хруста-

*) Более полное описание оптических приборов можно найти, например, в [1—4].

**) Для детального знакомства с физиологической оптикой мы отсылаем читателя к классическому труду Гельмгольца [5] и к книге Соуталла [6].

лик имеет форму двояковыпуклой линзы с плотностью (а следовательно, и показателем преломления), снижающейся по направлению к периферии, что уменьшает создаваемую им сферическую аберрацию (см. § 5.3). Средняя величина показателя преломления хрусталика равна 1,44, тогда как у *стекловидного тела*, находящегося между хрусталиком и сетчаткой, она составляет 1,34, т. е. близка к показателю преломления воды.

Радиус кривизны хрусталика может изменяться сокращением мышц, меняя таким образом фокусировку глаза. Эта способность называется *аккомодацией*. Наименьшее расстояние отчетливого видения для нормального, или *эметропического*, глаза примерно равно 25 *см*. Аккомодация глаза в большей степени изменяется от одного индивидуума к другому; она изменяется также с возрастом *).

Исправление значительных недостатков глаз с помощью дополнительных линз (очков) составляет предмет *офтальмологии*. Глазу свойственны три основных недостатка.

а. *Миопия*, или *близорукость*, при которой лучи от бесконечно удаленного точечного источника фокусируются перед сетчаткой. Она корректируется рассеивающей линзой, помещаемой перед глазом.

б. *Гиперметропия*, или *дальнозоркость*, при которой истинный фокус лучей от бесконечно удаленного предмета лежит за сетчаткой. Она корректируется собирательной линзой.

в. *Астигматизм*, при котором преломляющая способность глаза различна в разных плоскостях, проходящих через его оптическую ось. Этот дефект корректируется цилиндрической или тороидальной очковой линзой.

Сетчатка содержит два основных типа светочувствительных приемников, *палочки* и *колбочки*. Последние, согласно принятой теории цветного зрения, делятся на три типа цветочувствительных элементов. Колбочки преобладают в средней части сетчатки и их концентрация наиболее велика в так называемой *центральной ямке желтого пятна*. Эта область сетчатки ограничивает поле наиболее отчетливого зрения размером, несколько меньшим 1°, и определяет *фиксацию*, или направление «видения», глаза. Диаметр колбочек в центральной ямке меняется от 0,0015 до 0,005 *мм*. Палочки, малочувствительные к цвету, доминируют в периферических участках сетчатки. Они особенно чувствительны к свету малоинтенсивных источников. С этим связано явление «бокового зрения», которое проявляется, например, когда мы рассматриваем картину ночного неба.

Существует некий конечный предел в отчетливости деталей, который может быть достигнут с любой оптической системой. Этот предел, как будет показано в п. 8.6.2, в конечном счете обусловлен волновой природой света. Большой интерес для конструкторов-оптиков представляет величина нижнего предела разрешения глазом двух соседних точечных объектов, или *острота зрения*, поскольку от этой величины зависят все допуски (как оптические, так и механические), приемлемые при создании приборов для визуальных наблюдений. Для нормального глаза предельное угловое разрешение составляет около 1′, что при фокусном расстоянии 15 *мм* соответствует 0,0045 *мм* на сетчатке. При диаметре зрачка, бóльшем 5 *мм*, хроматическая и сферическая аберрации ухудшают разрешающую способность глаза. Поэтому обычно при конструировании приборов для визуальных наблюдений предполагается, что диаметр светового пучка, попадающего в глаз, не превышает 4—5 *мм*. При расчете таких приборов почти никогда не учитываются недостатки глаза, так как они меняются от человека к человеку.

---

*) К 10 годам оптическая сила глаза может изменяться в пределах 14 диоптрий, к 25 годам — 10 диоптрий и к 50 годам — 2,5 диоптрий.

§ 6.2. Фотографический аппарат

В *фотографическом аппарате* (рис. 6.2) действительное обратное изображение предмета образуется одной линзой (или комбинацией линз) на поверхности фотографической пленки или пластинки. Предмет может быть неподвижным или движущимся относительно камеры. В последнем случае необходимы короткие выдержки и соответственно большая апертура, обеспечивающая достаточное количество света. Как следует из (4.8.24), количество света от протяженного предмета, приходящееся на единицу площади изображения (освещенность E), пропорционально отношению d^2/f^2, где d — диаметр входного зрачка, а f — фокусное расстояние, т. е. обратно пропорционально квадрату числа F (номинальное фокальное отношение). Обычно его можно менять с помощью переменной *диафрагмы*.

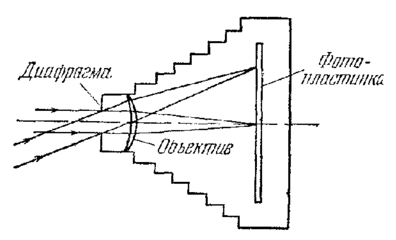


Рис. 6.2. Фотографический аппарат.

В отсутствие аберраций линейный размер изображения удаленного объекта, видимого под углом $\delta\theta$ в первой узловой точке, равен $f\delta\theta$. Следовательно, для получения изображения большого размера фокусное расстояние должно быть большим. Таким образом, для фотоаппарата важны две основные характеристики: фокусное расстояние его оптической системы и диапазон изменения фокального отношения, который эта система допускает. Еще одно требование, которое можно предъявить фотоаппарату — это большая широкоугольность. Фотоаппараты часто конструируют так, чтобы поле их зрения охватывало углы от 10 до 90°, а фокусные расстояния варьировали от 25 *мм* до 2 *м*. Еще большие фокусные расстояния (2—20 *м*) используются в астрономической фотографии, где поле очень малых угловых размеров ($\sim 1°$) с точечными объектами часто должно изображаться в большом масштабе.

В простых небольших фотоаппаратах используется одиночная собирательная линза, обычно в виде мениска (вогнуто-выпуклая), с диафрагмой, несколько

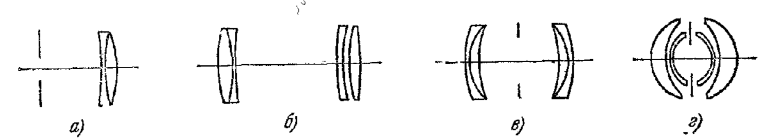


Рис. 6.3. Объективы разного типа.

а — объектив Шевалье, *б* — объектив Петцваля, *в* — объектив Рапид Ректилинеар; *г* — объектив Топогон.

отодвинутой от нее, как это показано на рис. 6.2. Такие линзы, не исправленные на хроматическую аберрацию, оставались вполне пригодными до тех пор, пока фотографические эмульсии были чувствительны главным образом к небольшому участку в синей части спектра (4000—4800 Å). Подобные линзы применялись в дешевых типах фотоаппаратов с фокальным отношением $f/11$ и с полем 45°.

Мениск, названный «*ландшафтной линзой*», появился около 1812 г., но вскоре его заменил ахроматический дублет Шевалье (рис. 6.3, *а*). Хотя с помощью ахроматической комбинации и удалось совместить фокусировку фотографически активного света и фокусировку всего видимого света (максимум яркости около $\lambda = 5800$ Å), однако при фокальном отношении, меньшем $f/16$, хорошего изображения получить не удавалось. Потребность в более свето-

сильных объективах (т. е. с меньшим фокальным отношением) для портретной фотографии привела к широкому распространению *портретного объектива Петцваля*, созданного около 1840 г. (рис. 6.3, *б*). Этот объектив состоял из двух отдельных дублетов. Кривизна поля изображения, создаваемая таким объективом, довольно значительна, но для поля, не превышающего нескольких градусов, достигается хорошая четкость при фокальном отношении до $f/4$. Улучшенный Даллмейером объектив Петцваля широко применяется в кинопроекторах.

Примерно с 1841 г. стало известно, что симметричные объективы всегда свободны от дисторсии, что строго справедливо только для случая единичного увеличения, но достаточно близко к истине и для других рабочих расстояний. Этот принцип лег в основу расчета целого ряда пряморисующих фотографических объективов. Объектив Рапид Ректилинеар Даллмейера (рис. 6.3, *в*), существующий с 1866 г., применялся во многих фотоаппаратах и в конце концов в 1920 г. его фокальное отношение стало меньше $f/8$, а поле составляло примерно 45°. Многие фирмы выпускали такие объективы и они получили общее название «*апланатов*». Тот же принцип симметрии, позволяющий устранить дисторсию, многократно использовался при конструировании широкоугольных объективов. Первый объектив такого типа *Гипергон Герца* 1900 г. мог охватить плоское поле размером до 150° при фокальном отношении около 25, хотя значительно чаще достаточным было поле в 90° при $f/16$. Четырехкомпонентный *Топогон* (рис. 6.3, *г*) 1933 г. с полем в 90° при $f/6{,}3$ лег в основу конструкции современных широкоугольных объективов.

В течение большой части девятнадцатого столетия конструкторы объективов были ограничены в своем выборе двумя основными типами стекол: боросиликатным кроном и тяжелым флинтом. Около 1890 г. Иенская компания ввела совершенно новый сорт стекла — бариевый крон, имеющий большой показатель преломления при малой дисперсии. Применение стекла такого типа в системе линз объектива позволило добиться большей степени коррекции. Рудольф из Иены, с чьим именем связано появление многих типов фотографических объективов, придумал название «*анастигмат*» для объектива, у которого для одного внеосевого направления астигматизм может быть полностью устранен. Пользуясь только старыми сортами стекол, Рудольф изготовил анастигмат, который позже стал известен под названием *Протар Цейсса* (рис. 6.4, *а*).

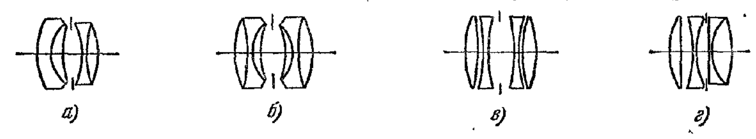


Рис. 6.4. Объективы разного типа.
а — объектив Протар; *б* — объектив Дагор; *в* — объектив Селор; *г* — объектив Тессар.

Применив же три сорта стекла, он смог изготовить анастигматический триплет. Два анастигматических триплета, соединенные симметрично, дали хорошо скорректированную систему. Такой объектив выпускался Цейссом под названием *Трипль Протар* и Герцем — под названием *Дагор* или *двойной анастигмат* (рис. 6.4, *б*). Типичный объектив подобного типа может охватить поле в 50° при $f/4{,}5$. Позднее Рудольф показал, что если каждая половина образована четырьмя склеенными компонентами, то она может быть «самостоятельной», т. е. половины с разными фокусными расстояниями можно использовать либо по отдельности, либо в комбинации. Такой объектив был изготовлен в 1894 г. и называется *Дубль Протар*.

Объектив, состоящий из четырех отдельных элементов, был усовершенствован Герцем в 1898 г. и получил название *Селор* (рис. 6.4, *в*); он стал прототипом

некоторых известных объективов с фокальным отношением до $f/4,5$ (например, *Унифокаль* Штейнхеля и *Авиар* Кука для аэрофотографии). В 1902 г. Рудольф создал свой *Тессар* (рис. 6.4, *г*), до сих пор пользующийся широкой известностью как относительно недорогой объектив, дающий превосходное изображение по полю до 60° с фокальным отношением, меньшим $f/3,5$. Изготовляют такие объективы с самыми разными фокусными расстояниями.

Старая конструкция Рудольфа — симметричный *Планар* 1895 г.— состояла из двух менисковых линз, служащих первым и последним элементами системы (рис. 6.5, *а*). После 1920 г. было разработано много очень светосильных

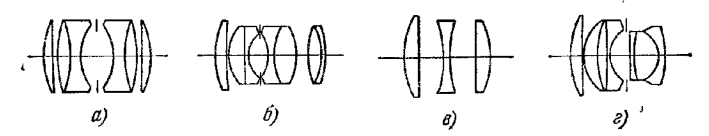


Рис. 6.5. Объективы разного типа.
а — объектив Планар; *б* — объектив Биотар Цейсса, *в* — триплет Кука, *г* — объектив Зоннар

объективов, у которых первым элементом была менисковая линза, а последним в большинстве случаев склеенный дублет (например, *Биотар Цейсса* (рис. 6.5, *б*), рассчитанный Мертье).

Другое направление в конструировании объективов, начатое в 1895 г. Тейлором, возникло благодаря усовершенствованию трехлинзового анастигмата, известного под названием *триплет Кука* (рис. 6.5, *в*). Достоинство такой трехэлементной системы заключается в возможности многочисленных вариаций параметров отдельных элементов. В недорогих фотоаппаратах эти «анастигматические» конструкции повсюду заменили «апланатические» объективы старых систем типа Рапид Ректилинеар. Светосильные объективы, предшественником которых можно считать триплет Кука (например, *Зоннар* $f/1,5$, рис. 6.5, *г*), широко используются в кинокамерах и миниатюрных фотоаппаратах. У Зоннара центральный отрицательный элемент и второй положительный — триплеты.

В камерах, предназначенных для специальных целей с большими эффективными фокусными расстояниями и соответственно малыми полями зрения, можно

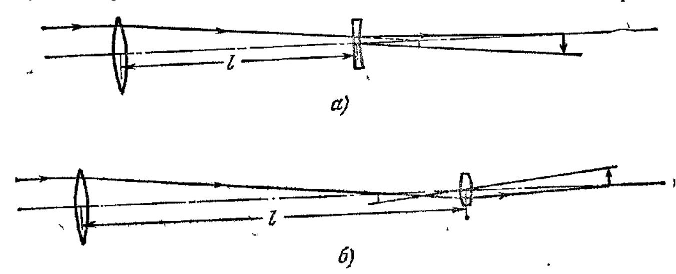


Рис. 6.6. Телеобъектив (*а*) и фотогелиограф (*б*).

использовать оптическую систему, состоящую из объектива и линзы, удлиняющей фокус. В *телеобъективах* удлинение фокуса обеспечивается рассеивающей линзой, помещенной перед первичным изображением (рис. 6.6, *а*). В *фотогелиографе* — приборе, предназначенном для фотографирования Солнца в большом масштабе, обычно применяется удлиняющая система собирающего типа (рис. 6.6, *б*). Эффективное фокусное расстояние такой системы приближенно

определяется выражением (4.7.3), т. е.

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}-\frac{l}{f_1 f_2},$$

где f_1 и f_2 — соответственно фокусные расстояния объектива и удлиняющей линзы, а l — расстояние между ними. У подобных систем эффективное фокусное расстояние может превышать расстояние от объектива до фокальной плоскости в 15 раз *).

В заключение следует отметить, что число F характеризует «скорость действия» фотографического объектива только в случае протяженного объекта. В случае точечного источника (что действительно имеет место в астрономии) свет в плоскости изображения в идеале будет концентрироваться на исчезающе малой площадке. Тогда лучшей характеристикой оптической силы объектива будет не число F, а квадрат диаметра апертуры. На самом же деле все обстоит значительно сложнее и существует ряд причин, увеличивающих изображение точечного источника до конечной (хотя часто и очень малой) величины. Главные из них — дифракция (см. гл. 8), зернистость фотографической эмульсии и нестабильность атмосферы.

§ 6.3. Линзовый телескоп **)

Телескоп — это оптическая система, с помощью которой можно рассматривать увеличенное изображение удаленного объекта. На рис. 6.7 схематически показана *астрономическая труба*, или *рефрактор*. Такой телескоп состоит из двух собирательных линз; первая, *объектив* — обычно ахроматический дублет,

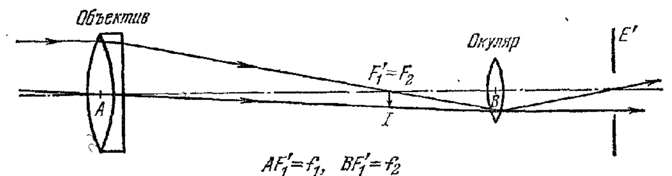


Рис. 6.7. Астрономическая труба, или рефрактор.

дает перевернутое изображение I, которое рассматривается с помощью второй линзы, или *окуляра*. В обычных условиях задняя фокальная плоскость объектива совпадает с передней фокальной плоскостью *окуляра*, так что падающий на объектив параллельный пучок лучей выходит из окуляра также параллельным. С помощью вспомогательной линзы перевернутое изображение можно превратить в прямое.

Входной зрачок системы совпадает с объективом; его изображение остальной частью системы, т. е. выходной зрачок, обозначен на рис. 6.7 через E'. Глаз должен располагаться так, чтобы его входной зрачок совмещался с E'; тогда весь свет, входящий в объектив под разными углами к оси, попадет в глаз.

Увеличение системы, применяемой для рассматривания объектов, находящихся в бесконечности, определяется как угловое увеличение на зрачках. По формуле Смита — Гельмгольца (4.4.49) оно равно величине, обратной линейному увеличению в плоскостях зрачков, и, значит, равно отношению радиусов входного и выходного зрачков. Выражение для увеличения получается непо-

*) Более полное описание фотографических объективов можно найти в книге [7]. Более поздние конструкции описаны в статьях [8—11].

**) Для более полного знакомства с телескопами см., например, [12].

средственно из (4.3.31), откуда, полагая, что углы малы, и заменяя $\operatorname{tg}\gamma$ и $\operatorname{tg}\gamma'$ на γ и γ', находим

$$\frac{\gamma'}{\gamma}=\frac{f_1}{f_2}.$$

Глаз можно аккомодировать на расстояние, отличное от бесконечности, а телескоп легко переюстировать так, чтобы он давал мнимое изображение на некотором расстоянии D от глаза *). Однако D не должно становиться меньше 25 *см*, что, как отмечалось выше, является ближайшим расстоянием, на котором нормальный глаз без напряжения отчетливо видит предмет.

Первый телескоп, о существовании которого достоверно известно, был телескоп Галилея, построенный в 1609 г. В этой системе (рис. 6.8) объективом служила собирательная линза, а окуляром — рассеивающая, помещенная перед

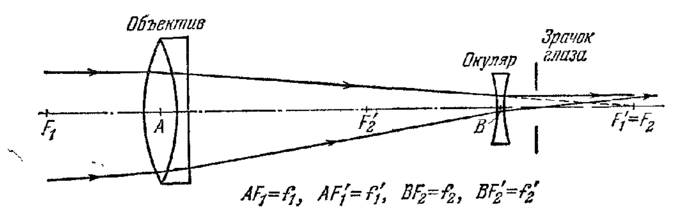


Рис. 6.8. Телескоп Галилея.

первичным изображением в таком месте, чтобы фокусы обеих линз совпадали. При таком расположении прямое изображение лежит в бесконечности, а промежуточное отсутствует. В телескопе подобного типа, известном под названием *телескопа Галилея*, апертурная диафрагма не совпадает с плоскостью объектива, поскольку изображение объектива окуляром находится между линзами и, следовательно, не доступно глазу. Глаз помещается непосредственно за окуляром, так что зрачок глаза становится апертурной диафрагмой и выходным зрачком. Главный луч для точек на границе поля пройдет очень близко от края объектива, и, следовательно, последний служит диафрагмой поля зрения, а не апертурной диафрагмой.

К недостаткам телескопа Галилея следует отнести малое поле зрения и отсутствие действительного изображения. Последнее обстоятельство не позволяет использовать в нем ни креста нитей, ни сетки. Вместе с тем малая длина телескопа, а также то обстоятельство, что он дает прямое изображение, делает его особенно удобным для театральных биноклей с увеличением от двух до трех раз.

В *бинокле* с такой же оптической системой, как и у астрономического телескопа, прямое изображение получается путем четырехкратного отражения (рис. 6.9). Последние происходят при падении света под углом 45° на поверхность стекло — воздух в так называемых *призмах Поро*. Этот угол, конечно, больше критического. Оборачивающая система такого рода удобна и часто применяется в небольших оптических приборах. В бинокле другого типа изображение переворачивается *призмой Кенига*. Эта сложная система с двумя взаимно перпендикулярными (с точностью до 1″) отражающими поверхностями, так называемой «крышей», ставится на пути светового пучка, как показано на рис. 6.10. Увеличенное расстояние между объективами у биноклей более распространенного типа полезно, так как усиливает стереоскопический эффект **).

*) При этом увеличение будет равно $\frac{\gamma_1}{\gamma}=\frac{f_1}{f_2'}+\frac{f_1}{D}$.

**) Хорошее описание биноклей можно найти в книге [13].

Земные зрительные трубы в основном подобны астрономическим телескопам, за исключением того, что изображение у них должно быть прямым. Для переворачивания изображения служат либо призмы, как в бинокле, либо дополнительные линзы. Типичная система такого рода показана на рис. 6.11.

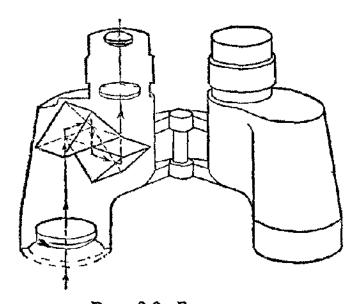


Рис. 6.9. Бинокль.

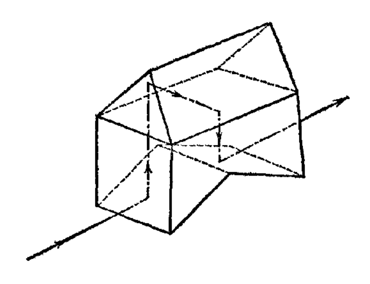


Рис. 6.10 Оборачивающая призма Кенига

В качестве объектива телескопа обычно применяют ахроматическую комбинацию двух линз, у которых радиусы кривизны выбраны так, чтобы сферическая аберрация была небольшой. Для этой цели подходит склеенный дублет

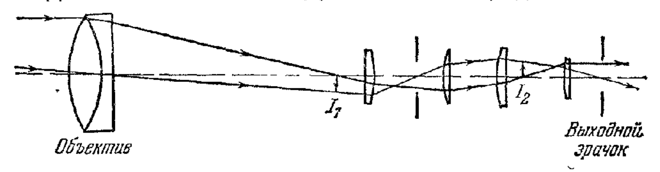


Рис. 6.11. Земная зрительная труба

с первой симметричной двояковыпуклой линзой из крона и второй плоско-вогнутой из флинта (рис. 6.12, *а*). Если нужно устранить не только сферическую аберрацию, но и кому (как, например в высококачественном астрономическом телескопе), то возникают большие трудности при расчете обеих несклеенных компонент объектива, от радиуса кривизны которых зависит апланатичность конструкции. Примерная конструкция такого типа с вогнуто-выпуклой компонентой из флинта показана на рис. 6.12, *б*.

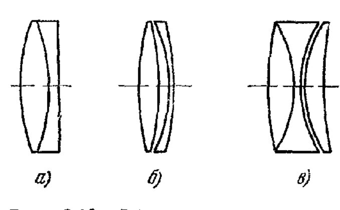


Рис. 6.12. Объективы различного типа.
а — склеенный ахромат *б* — апланатический ахромат, *в* — фотовизуальный объектив

Чрезмерную вогнутость первой поверхности можно уменьшить, если в свободную от комы конструкцию объектива ввести несферическую поверхность, дающую на оси системы стигматическое изображение.

Наиболее распространенный тип объективов с передней линзой из крона — это *объектив Фраунгофера*. *Объектив Штейнхеля* с передней линзой из флинта менее известен вследствие повышенной чувствительности флинта (по сравнению с кроном) к воздействию атмосферы.

Фотовизуальный объектив — третий тип объектива, широко применяемый в рефракторах, показан на рис. 6.12, *в* Использование трех различных сортов стекла позволило значительно уменьшить хроматическую разницу фокусных

расстояний в заданном спектральном интервале. Такой объектив почти одинаково хорош как для фотографических, так и для визуальных наблюдений и одно время был широко распространен в астрономических инструментах с диаметрами отверстий до 30 *см*. Значительная кривизна центральной компоненты, затрудняющая центрировку системы, и нестабильность флинтовой компоненты уменьшают, однако, ценность этих объективов.

Окуляры зрительных труб, вообще говоря, должны удовлетворять двум требованиям. Во-первых, фокусное расстояние такого окуляра должно обеспечивать необходимое увеличение и, во-вторых, его апертура должна быть достаточной, чтобы собрать свет от протяженного изображения. Этим требованиям удовлетворяет одиночная линза (рис. 6.13, *а*), но более удобна система из двух меньших линз, показанная на рис. 6.13, *б*, где изображена оптическая схема *окуляра Рамсдена*.

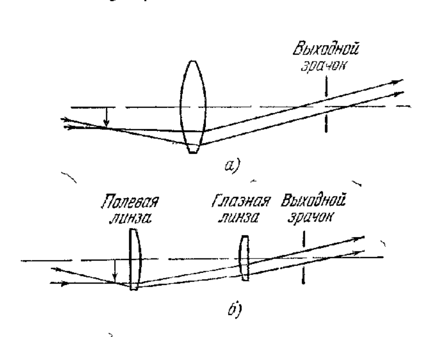


Рис. 6.13. Однолинзовый окуляр (*а*) и окуляр Рамсдена (*б*)

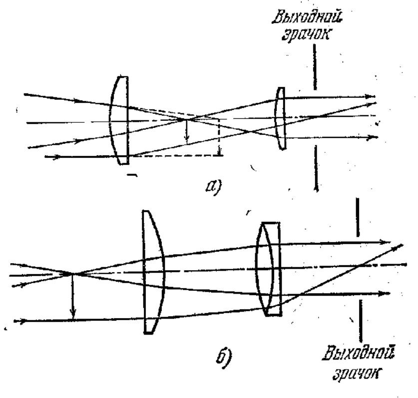


Рис. 6.14. Окуляр Гюйгенса (*а*) и окуляр Келлнера (*б*).

Окуляр должен давать изображение апертурной диафрагмы в таком месте, где удобно поместить глаз. Расстояние между этим изображением и последней поверхностью окуляра называется *удалением выходного зрачка*.

Первая линза двухкомпонентного окуляра называется *полевой линзой*, так как ее часто помещают очень близко от первичного изображения. Она отделена от него частично для того, чтобы стало незаметным любое загрязнение или дефект линзы, и частично для того, чтобы в плоскость изображения можно было поместить крест нитей или сетку. Вторая компонента окуляра называется *глазной линзой*.

Наиболее широкое распространение получили окуляр Рамсдена и изображенный на рис. 6.14, *а* окуляр Гюйгенса. Полевая линза у последнего находится впереди первичного изображения. Недостатком окуляра Гюйгенса является небольшое удаление выходного зрачка; кроме того, им нельзя пользоваться с внешним крестом нитей. В этом окуляре сферическую аберрацию нельзя исправить так же хорошо, как в окуляре Рамсдена, но зато он свободен от поперечной хроматической аберрации и внеосевой комы.

Применение двух линз вместо одной (см. рис. 6.14) способствует устранению хроматической аберрации, и коррекцию окуляра Рамсдена можно значительно улучшить путем замены одиночной глазной линзы дублетом. Еще более сложны «ортоскопические» и «симметричные» окуляры, которые в основном подобны окуляру Рамсдена, но состоят из четырех компонент. Они в высокой степени свободны от аберраций (особенно от дисторсии) и обеспечивают значительное удаление выходного зрачка при данном увеличении.

Принцип устройства *перископа*, т. е. прибора, предназначенного для рассматривания предметов, расположенных так, что прямому наблюдению их мешает какое-либо препятствие, тот же что и у зрительной трубы. Однако два дополнительных требования значительно усложняют его конструкцию. Во-первых, в перископе ход лучей должен иметь вид ломаных линий, а сам он должен давать прямое, неперевернутое изображение. Для получения такого изображения необходимо либо четное число отражений, либо полное их отсутствие. На рис. 6.15 показаны два простых устройства; в одном из них входящий и выходящий световые пучки распространяются в одну сторону, в другом направления пучков противоположны. Первое должно работать совместно со зрительной трубой, дающей прямое изображение, второе — с оборачивающей трубой. Еще одно устройство показано на рис. 6.16, направление выходящего из него

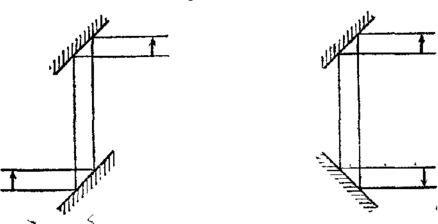


Рис. 6.15. Перископические зеркала.

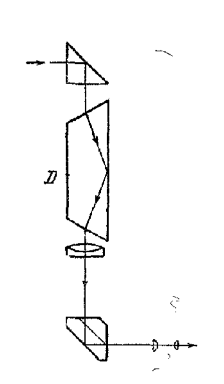


Рис. 6.16. Перископ с призмой Дове

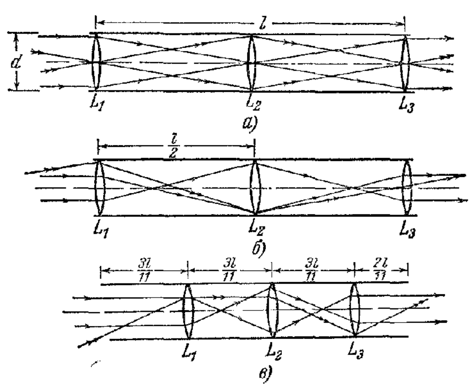


Рис. 6.17. Трехлинзовые системы с единичным увеличением

пучка остается постоянным, тогда как вращение первой призмы обеспечивает возможность наблюдения в разных направлениях. Прямое изображение в этой системе сохраняется с помощью *призмы Дове D*, которая дает одно дополнительное отражение и вращается со скоростью в два раза меньшей, чем первая приз-

Таблица 6.1

	Фокусные расстояния линз			Изображение	Апертура на оси	Поле (рад)	Виньетирование на краю поля
	L_1	L_2	L_3				
а	$l/2$	$l/4$	$l/2$	Обратное	d	$2d/l$	0
б	$l/6$	$l/6$	$l/6$	Прямое	$d/2$	$2d/l$	100%
в	$l/11$	$l/11$	$l/11$	Прямое	$d/2$	$11d/2l$	~86%

ма. Нижняя «крышеобразная» призма (Амичи) дополняет число отражений до четырех.

Второе требование, предъявляемое к перископу, заключается в том, что он должен иметь большое угловое поле зрения, хотя сама перископическая система заключена в узкую трубу. Труба диаметром 20 *см* и длиной 12 *м* допускает максимальное поле зрения, равное лишь 2°. Такие требования можно удовлетворить, только заставив свет проходить вдоль трубы сквозь ряд линз. Расстояния между ними могут быть различными. На рис. 6.17 показаны три трехлинзовые системы с единичным увеличением, а в табл. 6.1 приведены величины полей зрения и эффективные апертуры этих систем. Очевидно, что с увеличением числа линз диапазон этих величин можно расширить *).

§ 6.4. Зеркальный телескоп

В *астрономических зеркальных телескопах*, или *рефлекторах*, свет от небесных объектов падает на вогнутое главное зеркало, выполняющее ту же роль, что и объектив рефрактора, а именно, образуя действительное изображение объекта в своей фокальной плоскости. Это изображение либо получается непосредственно на фотографической пластинке, либо исследуется визуально через окуляр. Астрономические телескопы подобного типа особенно широко используются в наши дни. Так же как и у рефрактора, увеличение такого телескопа равно отношению фокусных расстояний объектива и окуляра.

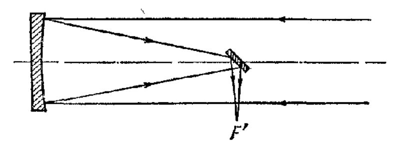


Рис. 6.18. Телескоп Ньютона.

Первый зеркальный телескоп был построен в 1668 г. Ньютоном. В созданном им телескопе на пути лучей, отраженных от главного зеркала, устанавливалось еще небольшое плоское зеркальце, отклоняющее лучи так, чтобы удобно было рассматривать изображение (рис. 6.18). Такое устройство известно под названием *телескопа Ньютона*.

Для получения стигматического осевого изображения зеркало должно быть параболическим. Однако у параболического зеркала внеосевые изображения искажаются сильной комой, в результате чего при больших относительных апертурах пригодное для работы поле зрения становится очень малым. При зеркале диаметром 90 *см* с фокальным отношением $f/6$ поле зрения составляет примерно 20′, тогда как у наибольшего современного телескопа Хале с главным зеркалом диаметром 5 *м* угловое поле зрения равно приблизительно 45′ [15] **).

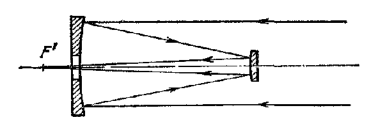


Рис. 6.19 Телескоп Кассегрена.

В других типах зеркальных телескопов главная оптическая система состоит из двух зеркал. Из них первое всегда вогнутое. В *телескопе Кассегрена* второе зеркало выпуклое (рис. 6.19). В больших астрономических телескопах такое устройство часто заменяет ньютоновское. Для получения разных фокусных расстояний можно использовать различные зеркала. Если в телескопах этой системы применяется параболическое главное зеркало, то второе зеркало должно быть гиперболическим. Можно выбрать и другие формы зеркал, дающие стигматическое осевое изображение при различных величинах внеосевой комы. В § 4.10 упоминалась двухзеркальная система, предложенная Шварцшильдом,

---

*) Дополнительные сведения можно найти в статье Смита [14].

**) Общее описание 5-метрового телескопа см. в статьях [16, 17].

которая полностью апланатична. Однако форма зеркал у нее более сложная. Система Кассегрена получила широкое распространение в земных зрительных трубах, предназначенных для рассматривания ландшафтов, так как при невысокой стоимости, компактности и большом фокусном расстоянии она дает большое увеличение, не уменьшая вместе с тем удаление выходного зрачка.

В менее известном *телескопе Грегори* вогнутое второе зеркало устанавливается позади фокуса главного зеркала.

У крупнейших рефлекторов с помощью второго зеркала легко получить очень большие фокусные расстояния, значительно превышающие фокусные расстояния наибольших рефракторов. В *системе Куде* свет отражается от третьего (плоского) зеркала вдоль оси вращения инструмента, что позволяет направить свет в неподвижный спектрограф.

Одно из существенных преимуществ зеркального телескопа — полное отсутствие хроматической аберрации. Поскольку кривизна зеркальной поверхности может быть значительно меньше, чем у линз объектива, можно строить зеркальные телескопы с фокальным отношением много меньшим, чем у рефракторов, что в свою очередь обеспечивает более яркое изображение и приводит к большей компактности оптической системы. Кроме того, зеркало можно сделать больше линзы, так как оптические неоднородности в толще стекла для зеркала не имеют значения.

Основными недостатками астрономических рефлекторов являются чувствительность зеркальной поверхности и самой трубы телескопа к изменениям температуры, малость углового поля зрения и трудность осуществления достаточно жесткой механической конструкции. Последнее обстоятельство служило причиной наибольших затруднений, возникавших у конструкторов телескопа Хале.

Успешное применение зеркального телескопа с фотографической регистрацией вызвало стремление увеличить поле зрения телескопа. Один из методов, пригодных для увеличения полезного поля, заключается в том, что непосредственно перед фотопластинкой устанавливается полевая линза, рассчитанная так, чтобы вносимая ей внеосевая кома, компенсировала кому главного зеркала. Таким путем с зеркалом диаметром 80 *см* достигается поле зрения порядка 1,5° [18].

Значительное увеличение углового поля можно получить при фотографировании с помощью большой зеркальной системы, предложенной в 1930 г. Шмидтом и упоминавшейся в § 4.10 [19, 20]. Параболическое зеркало дает безупречное изображение в осевых лучах и плохое изображение из-за комы на небольших расстояниях от оси. Сферическое зеркало с апертурной диафрагмой в центре кривизны C дает однородные изображения с сильной сферической аберрацией на широкой тоже сферической поверхности с тем же центром кривизны. Шмидт поместил в плоскости апертурной диафрагмы тонкую почти плоскопараллельную пластинку. Одна ее сторона была плоской, другая имела

Сферическое зеркало
Корректирующая пластинка
Диафрагма

Рис. 6.20. Камера Шмидта.
Кривизна корректирующей пластинки значительно преувеличена

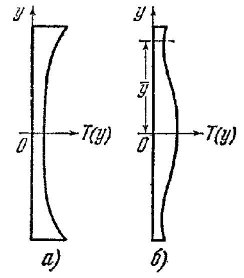


Рис. 6.21. Сечение корректирующих пластинок Шмидта.
Кривизны значительно преувеличены

сложный профиль (рис. 6.20, 6.21). Такая пластинка называется *корректирующей*, и ее назначение заключается в предварительной коррекции входящего в систему плоского волнового фронта с целью полной компенсации сферической аберрации, вносимой сферическим зеркалом.

Нетрудно вывести уравнение для асферической поверхности корректирующей пластинки *камеры Шмидта*. Для этого сравним уравнение сферического и параболического зеркал. Уравнение сферического зеркала радиуса $R = 2f$ имеет вид (рис. 6.22)

$$(z^{(s)}-2f)^2+y^2=(2f)^2,$$

т. е.

$$z^{(s)}=\frac{y^2}{4f}+\frac{y^4}{64f^3}+\ldots,$$

тогда как для параболического зеркала того же параксиального радиуса ($R = 2f$) имеем

$$z^{(p)}=\frac{y^2}{4f}.$$

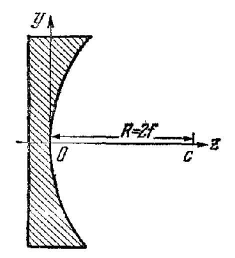


Рис. 6.22 Сферическое зеркало.

Поскольку пучок параллельных лучей, идущих в направлении оптической оси, при отражении от параболического зеркала остается стигматическим, в зоне радиуса y величина предварительной коррекции, которую нужно внести в волновые фронты, падающие на сферическое зеркало, приблизительно равна

$$2(z^{(s)}-z^{(p)})=\frac{y^4}{32f^3}.$$

Мы пренебрегли здесь членами с y в шестой и более высоких степенях *). Пройдя корректирующую пластинку, лучи останутся почти параллельными, так что практически не важно, измеряется ли y на зеркале или на пластинке. Следовательно, если n — показатель преломления пластинки, то ее толщина $T(y)$ в зоне радиуса y должна превышать ее толщину на оси $T(0)$ на величину

$$(n-1)[T(y)-T(0)]=2(z^{(s)}-z^{(p)}),$$

т. е.

$$T(y)-T(0)=\frac{y^4}{32(n-1)f^3}. \qquad (1)$$

Сечение такой корректирующей пластинки показано на рис. 6.21, *а*.

Необходимое отклонение от сферичности очень невелико. Для примера рассмотрим камеру Шмидта $f/3{,}5$ с апертурным диаметром корректирующей пластинки $2y_0=40$ *см* и, следовательно, фокусным расстоянием $f=140$ *см*. При $n=1{,}5$ максимальная асферичность $[T(y)-T(0)]_{\text{макс}}$, согласно (1), равна

$$\frac{y^4}{32(n-1)f^3}=0{,}0036\ \textit{см}.$$

Дальнейшего улучшения можно добиться сравнением сферического зеркала с несколько иным «сравнительным параболоидом». Если f' — фокусное расстояние параболоида, то $z^{(p)}=y^2/4f'$, и вместо (1) для поверхности пластинки получим

$$T(y)-T(0)=\frac{y^4-Ay^2}{32(n-1)f^3}, \qquad (2)$$

*) Можно показать, что это приближение допустимо, если фокальное отношение для камеры Шмидта не меньше чем приблизительно $f/3$.

где

$$A = 16f^3 \left(\frac{1}{f'} - \frac{1}{f}\right).$$

С корректирующей пластинкой (2) лучи, параллельные оси, сойдутся после прохождения системы в точке, находящейся на расстоянии f' от осевой точки зеркала. Эту добавочную степень свободы (выбор величины f') можно использовать для уменьшения хроматической аберрации, вносимой пластинкой. Так как

$$\frac{\partial T}{\partial y} = 0, \quad \text{когда} \quad y(2y^2 - A) = 0,$$

то луч, параллельный оси и отстоящий от нее на расстоянии $y = \sqrt{A/2}$, не отклоняется пластинкой. Зона радиуса $y = \sqrt{A/2}$ называется *нейтральной зоной*. Для того чтобы хроматическая аберрация стала минимальной, нейтральная зона должна доходить почти до края пластинки, как это показано на рис. 6.21, *б* Точнее, можно показать, что диаметр кружка смешанных цветов будет наименьшим, если радиус нейтральной зоны составляет приблизительно 0,87 радиуса пластинки. Это соответствует выбору $A = {}^3/_4 y_0^2$ [21, 22].

Конечно, корректирующая пластинка отклоняет проходящие через нее косые пучки иначе, чем пучки, падающие на нее нормально; поэтому такая система не свободна от внеосевых аберраций. Следует подчеркнуть, что качество

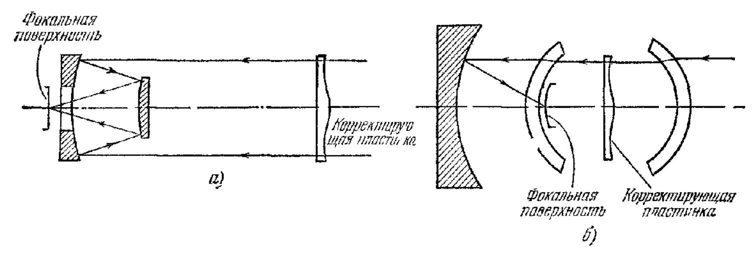


Рис. 6.23. Камера Шмидта — Кассегрена (*а*) и суперкамера Бекера — Шмидта (*б*)

изображений, полученных на сферической поверхности, концентрической главному зеркалу, превосходно *). Это относится также и к камере Шмидта с малым фокальным отношением, работающей по полю, во много раз превышающему поля, получающиеся с обычными системами.

Рассматривались и более сложные системы, использующие принцип камеры Шмидта **), а некоторые из них были даже построены Наиболее интересны: *камера Шмидта — Кассегрена*, в которой используются два зеркала (сферические или несферические) совместно с корректирующей пластинкой (рис. 6.23, *а*), и *суперкамера Бекера — Шмидта*, состоящая из сферического зеркала, двух менисков и корректирующей пластинки (рис. 6.23, *б*). К достоинствам первой следует отнести плоское поле изображения, более удобное, чем в камере Шмидта, а также то, что ее общая длина меньше, чем у камеры Шмидта с тем же фокусным расстоянием. Суперкамера Бекера — Шмидта может работать с предельно малыми фокальными отношениями.

---

*) Это отчасти связано с тем, что кривизна поля, как можно показать, является единственной первичной аберрацией [21] Аберрации пятого порядка в камере Шмидта исследовались Каратеодори [23] и Линфутом [22]

**) Обзор таких систем см. в [24—26] Более полное описание суперкамеры Бекера—Шмидта см. [27]

§ 6.5. Осветители

Назначение осветителей — собрать как можно больше света от искусственного источника и направить его с допустимым углом расхождения во входной зрачок оптической системы. Источники света имеют конечные размеры, и поэтому конструкция осветительного прибора в такой же степени зависит от природы источника, как и от назначения требуемого освещения. На рис. 6.24 показан входной зрачок оптической системы, которая требует освещения по всем

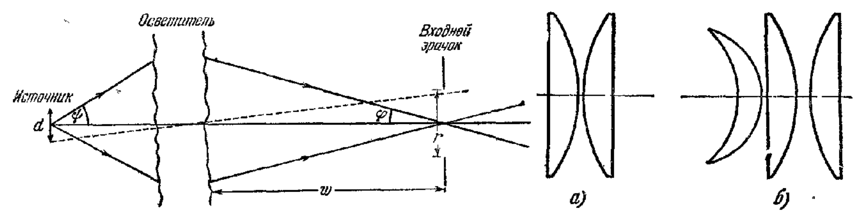


Рис. 6.24. К работе осветительного прибора.

Рис. 6.25. Конденсоры.

направлениям, образующим с осью углы, не превышающие значения φ. Диаметр зрачка равен r. Пусть d — диаметр источника, который мы считаем круглым. Для большей эффективности осветительной системы она должна давать увеличение r/d, а ее апертура должна быть видна из центра входного зрачка под телесным углом 2φ. Рабочее расстояние w должно превышать некоторую заданную величину. Из рассмотрения рис. 6.24 следует, что чем меньше d, тем больше *угол охвата* 2ψ и, следовательно, потери света малы. Большие углы освещения (большое φ) часто позволяют уменьшить отношение w/r, что ведет к более эффективному использованию больших источников. Таким образом очевидно, что в каждом частном случае существуют ограничения на величину источника, который еще можно удовлетворительно использовать, и эти ограничения становятся менее строгими при больших допустимых углах освещения. В фотоувеличителе ($\varphi \approx 20°$) большие источники (матированные лампы или трубки с холодным катодом) часто обеспечивают хорошее освещение. Для кинопроектора ($\varphi \approx 5°$), от которого требуется максимальная освещенность, пригодна нить электролампы. Если размеры прожектора, который должен обеспечить большой поток в угле 1—2°, не чрезмерно велики, то в нем могут применяться только очень яркие источники малых размеров.

Очень часто осветитель может быть сконструирован просто из подходящего источника и конденсорной линзы. Для успешного использования малых источников эта линза должна быть по возможности свободна от аберраций. Каждый конденсор можно заменить «фокальной сферой», соответствующей источнику минимального размера, с которым еще можно эффективно работать. Конденсор из двух плоско-выпуклых линз (рис. 6.25, *а*) дает обычно вполне удовлетворительные результаты. Иногда предпочтение отдается более сложной трехлинзовой системе (рис. 6.25, *б*), тогда как в других случаях оказывается достаточной одиночная линза.

Условия освещения в микроскопах требуют более полного обсуждения и будут рассмотрены ниже, в пп. 8.6.3 и 10.5.2.

§ 6.6. Микроскоп

Кажущаяся величина предмета определяется величиной его изображения на сетчатке. В случае невооруженного глаза кажущийся размер зависит от угла, под которым предмет виден. Для нормального глаза наименьшее расстояние отчетливого зрения, как указывалось в § 6.1, примерно равно 25 *см*. Это рас-

стояние наиболее удобно для рассматривания деталей предмета. Если перед глазом поместить собирательную линзу, то рассматриваемый предмет можно значительно приблизить к глазу. И хотя такая линза образует мнимое увеличенное изображение на расстоянии, превышающем расстояние до предмета (рис. 6.26), рассматривать его удобнее, чем сам предмет.

Рис. 6.26. Лупа.

Увеличение приборов, служащих для рассматривания близких предметов, определяется как отношение углового размера изображения, находящегося на расстоянии наилучшего зрения (обычно 25 *см*) к угловому размеру предмета, отнесенного на то же расстояние. Это эквивалентно определению обычного линейного увеличения M в том случае, когда изображение находится на расстоянии наилучшего зрения от выходного зрачка прибора. Для простой линзы (см. рис. 6.26) имеем

$$M=\frac{Y'}{Y}=\frac{\zeta'}{\zeta}, \tag{1}$$

где $\zeta'=25$ *см* — расстояние наилучшего зрения. Согласно (4.4.31)

$$\frac{1}{\zeta}-\frac{1}{\zeta'}=\frac{1}{|f|}, \tag{2}$$

так что

$$M=1+\frac{\zeta'}{|f|}=1+\frac{25}{|f|}; \tag{3}$$

в последнем выражении фокусное расстояние f измеряется в сантиметрах. Так как f обычно мало по сравнению с 25 *см*, увеличение можно записать как

$$M\sim\frac{25}{|f|}. \tag{4}$$

На практике различные типы луп, например простая двояковыпуклая линза и ахроматический дублет, очень часто используются в карманных лупах

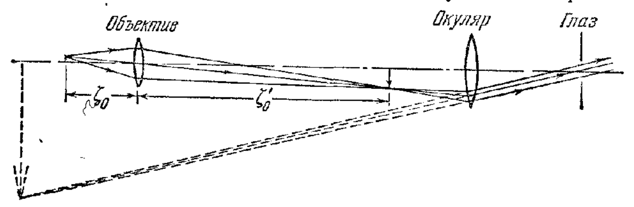


Рис. 6.27. Схема, иллюстрирующая принцип устройства микроскопа.

или в лупе часовщика. Часто встречается система из двух раздельных плосковыпуклых линз, сходная с окуляром Рамсдена (см. рис. 6.13, *б*). Очевидно что при большом поле зрения и большом увеличении расстояние между предметом и глазом становится недопустимо малым. По этой причине Галилей предложил около 1610 г. конструкцию *микроскопа*, оптическая система которого состояла из короткофокусного объектива и окуляра-лупы. В такой системе увеличение осуществляется в две ступени (рис. 6.27).

Объектив микроскопа образует увеличение изображения предмета в плоскости, удобной для рассматривания через *окуляр*. Увеличение объектива равно $-\zeta_0'/\zeta_0$, где ζ_0 и ζ_0' — рабочие расстояния объектива, удовлетворяющие соотношению (2). Если f_1 — фокусное расстояние окуляра, измеренное в сантиметрах, то его увеличение, согласно (4), равняется $25/|f_1|$. Следовательно,

увеличение всей системы равно

$$M = -\frac{25}{|f_1|} \cdot \frac{\zeta_0'}{\zeta_0}.$$

Отрицательная величина M указывает на то, что изображение получается перевернутым. Увеличения объектива и окуляра обычно даются в описании микроскопа отдельно.

Другая величина, представляющая интерес при описании микроскопического объектива,— это его числовая апертура (Ч. А.). Она была определена в п. 4.8.2 как произведение $n \sin \theta$, где θ — половина угловой апертуры в пространстве предмета (т. е. половина угла при вершине конуса лучей, исходящих из осевой точки предмета и попадающих на объектив), а n — показатель преломления среды в пространстве предмета. Эта величина служит мерой не только светосилы объектива, но, как будет показано в п. 8.6.3, и его разрешающей способности, т. е. предельной четкости деталей, которую он дает. Здесь мы расскажем о способах получения наивысшей числовой апертуры.

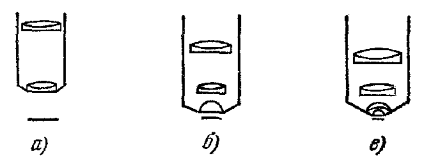


Рис 6.28 Объективы микроскопов.
а — малое увеличение, *б* — среднее увеличение; *в* — большое увеличение.

Объективы микроскопа, вообще говоря, должны быть хорошо исправлены на сферическую и хроматическую аберрации, а также на кому, так как они предназначены для работы с предельно большими апертурами. Для объектива с небольшим увеличением (~10×) можно воспользоваться системой из двух отдельных склеенных ахроматических линз (рис. 6.28, *а*), исправленной на сферическую аберрацию и кому. Однако такая комбинация линз не годится для объективов с большим увеличением (~50× и больше). В этом случае применяются другие системы линз, в которых используется существование на сферических поверхностях апланатических точек (см. п. 4.2.3). Это осуществляется

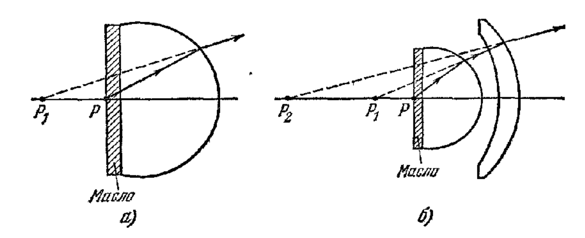


Рис. 6.29. Уменьшение углового расхождения лучей в микроскопе.

следующим образом. Предмет помещается в точке P вблизи плоско-выпуклой линзы (рис. 6.29, *а*). Пространство между предметом и линзой заполняется маслом с показателем преломления n, очень близким к показателю преломления линзы. Если C — центр кривизны исправленной поверхности линзы, а r — ее радиус кривизны, то мнимое изображение P_1 точки P, отстоящей от C на расстоянии r_1/n_1, находится на расстоянии $n_1 r_1$ от C и обе эти точки связаны условием апланатизма. Таким путем достигается большая числовая апертура и в системе уменьшается расхождение лучей, исходящих из P, без внесения монохроматических аберраций. Однако при этом возникает хроматическая аберрация и ее следует скомпенсировать остальной частью оптической системы.

Можно еще уменьшить расхождение лучей дополнительным положительным мениском *) (рис. 6.29, *б*). Центр кривизны передней поверхности мениска находится в точке P_1, а радиус его задней поверхности выбирается так, чтобы P была апланатической точкой этой поверхности. Лучи, преломленные на задней поверхности мениска, образуют мнимое изображение в другой апланатической точке P_2.

Добавление других менисков дает последовательный ряд мнимых изображений P_3, P_4, ..., лежащих все дальше и дальше от P. Таким образом, угловое расхождение лучей становится все меньше и меньше. Однако редко применяют больше двух линз, так как нельзя полностью скомпенсировать вносимую ими хроматическую аберрацию.

Описанные выше *иммерсионные объективы* позволяют работать с более широким конусом лучей, чем *сухие объективы* такого же диаметра. В случае сухих объективов лучи, выходящие из покровного стекла, защищающего рассматриваемый предмет, попадают в воздух и, преломляясь, отклоняются наружу. Следовательно, расхождение лучей увеличивается. При использовании же объективов с масляной иммерсией лучи, выходящие из покровного стекла, не преломляются. Наибольшую числовую апертуру (~1,4) можно получить только у таких объективов.

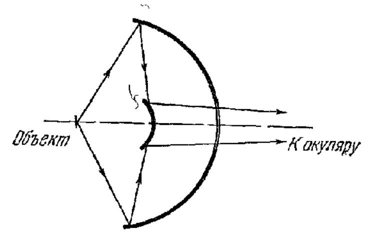


Рис. 6.30. Зеркальный объектив микроскопа Бурха.

Значительно большие требования предъявляются к объективам микроскопа, предназначенным для работы в ближней ультрафиолетовой ($\lambda \sim 2500$ Å), видимой и близкой инфракрасной областях спектра. Так, например, если микроскоп должен использоваться в сочетании со спектрографом и фотометром, необходима высокая степень ахроматизации.

Расчетом ахроматических объективов к микроскопам впервые систематически занялся Бурх [29] **), взявший за основу аналитическое решение Шварцшильда для двухзеркальной апланатической системы, о которой упоминалось в § 4.10. Бурх нашел, что если числовая апертура превышает 0,5, то по крайней мере одно из зеркал должно быть асферическим. Тем не менее в зеркальных системах невозможно осуществить такие же большие числовые апертуры, как в хороших иммерсионных объективах. Типичный ахроматический объектив Бурха показан схематически на рис. 6.30. Не считая ахроматичности, его основное достоинство заключается в большом рабочем расстоянии, т. е. расстоянии между предметом и ближайшей поверхностью объектива. Оно может быть столь же большим, как и фокусное расстояние объектива. Недостатком такой системы является присутствие второго зеркала, загораживающего часть света. В системах с малой числовой апертурой, использующих сферические поверхности, такое «загораживание» может превышать 45%.

Рассматривались также и зеркально-линзовые объективы микроскопов. Так, была предложена система, в которой отражения происходят внутри стекла на концентрических сферических покрытых серебром поверхностях [30]. Серебро на второй поверхности нанесено так, что «загораживание» уменьшено до 20%. Двухзеркальный объектив с относительно малым «загораживанием» описан Грейем [31]. Он применил кварцевые и флуоритовые компоненты с большой кривизной, которые при условии обеспечения необходимой «цветовой»

*) Этот метод предложен Амичи [28].

**) В первой его статье содержится также обзор ранних исследований по ахроматическим объективам для микроскопов.

коррекции в ультрафиолетовых и видимых областях спектра вполне заменяют поверхности сложной формы, необходимые при работе с зеркалами.

Обычно предметы, исследуемые под микроскопом, сами не светятся и, следовательно, нуждаются в постороннем освещении. Во многих случаях рассматриваемые предметы представляют собой тонкий срез прозрачного вещества и наблюдаются в проходящем свете. В системах с небольшой числовой апертурой (Ч. А. до 0,25) вполне достаточно рассеянного дневного света, отраженного под углом от вогнутого зеркала. В других случаях необходимо пользоваться искусственными источниками. Для большей концентрации света применяется вспомогательная система линз, или конденсор. Существует много различных способов освещения микроскопических объектов; два из них будут описаны в гл. 10.

В настоящем параграфе содержались только основные сведения о микроскопе. Полное обсуждение вопросов, связанных с формированием изображения в микроскопе, нуждается в более тонких методах (см. п. 8.6.3 и §§ 9.5 и 10.5).

ЛИТЕРАТУРА

1. Dictionary of Applied Physics, ed. by Glazebrook, Macmillan, London, 1923, Vol. IV.
2. L. C. Martin, An Introduction to Applied Optics, Pitman, London, 1950, Vol. II.
3. G. A. Boutry, Optique Instrumentale, Masson, Paris, 1946.
4. D. A. Jacobs, Fundamentals of Optical Engineering, McGraw-Hill, New York, 1943.
5. H. Helmholtz, Handbuch der Physiologischen Optik, L. Voss, Hamburg, Leipzig, 3 Aufl., I (1909), II (1911), III (1910).
6. J. P. C. Southall, An Introduction to Physiological Optics, Oxford Univ. Press, 1937.
7. W. Merté, R. Richter, M. Rohr, Das Photographische Objectiv, Springer, Vienna, 1932.
8. W. Taylor, H. W. Lee, Proc. Phys. Soc. 47, 502 (1935).
9. H. W. Lee, в сб. «Rep. Progr. Phys.», Phys. Soc. 1940, Vol. 7, p. 130.
10. R. Kingslake, в сб. «Proc. London Conf. on Opt. Instrum.», Chapman a. Hall, London, 1951, p. 1.
11. C. G. Wynne, в сб. «Rep. Progr. Phys.» Phys. Soc. 1956, Vol. 19, p. 298.
12. A. Danjon, A. Couder, Lunettes et Télescopes, Rev. d'Opt., Paris 1935.
13. M. Rohr, Die Binokularen Instrumente, J. Springer, Berlin, 1920.
14. T. Smith, в кн. «Dictionary of Applied Physics», ed. by R. Glarebrook, Macmillan, London, 1923, Vol. IV, p. 350.
15. F. E. Ross, Publ. Astr. Soc. Pac. 46, 342 (1934).
16. J. A. Anderson, Publ. Astr. Soc. Pac. 60, 221 (1948).
17. B. Rule, Publ. Astr. Soc. Pac. 60, 225 (1948).
18. F. E. Ross, Astrophys, J. 81, 156 (1935).
19. B. Schmidt, Central-Zeitung f. Optik u. Mechanik 52, Heft 2 (1935); Mitt. Hamburg. Sternw. Bergedorf 7, № 36, 15 (1932).
20. R. Schorr, Z. f. Instrkde 56, 336 (1936); Mitt. Hamburg. Sternw. Bergedorf 7, № 42, 175 (1936); Astr. Nachr. 258, 45 (1936).
B. Strömgren, Vierteljahrsschrift der Astr. Ges. 70, 89 (1935).
22. E. H. Linfoot, Mo. Not. Roy. Astr. Soc. 109, 279 (1949); в сб. «Recent Advances in Optics», Clarendon Press, Oxford, 1955, пп. 182, 192.
23. C. Carathéodory, Elementare Theorie des Spiegelteleskops von B. Schmidt, Teubner, Leipzig, 1940.
24. H. Slevogt, Z. f. Instrkde 62, 312 (1942).
25. H. Köhler, Astr. Nachr. 278, 1 (1949).
26. E. H. Linfoot, Mo. Not. Roy. Astr. Soc. 108, 81 (1948); в сб. «Recent Advances in Optics», Clarendon Press, Oxford, 1955, Ch. IV.
27. F. L. Whipple, Sky a. Telescope 8, 90 (1949).
28. G. B. Amici, Ann. de Chim. et Phys. 12, 117 (1844).
29. C. R. Burch, Proc. Phys. Sos. 59, 41, 47 (1947).
30. A. Bouwers, Achievements in Optics, Elsevier, Amsterdam, 1946.
31. D. S. Grey, в сб. «Proc. London Conf. on Opt. Instrum», Chapman a. Hall, London 1951, p. 65.

ГЛАВА 7

ЭЛЕМЕНТЫ ТЕОРИИ ИНТЕРФЕРЕНЦИИ И ИНТЕРФЕРОМЕТРЫ

§ 7.1. Введение

В гл. 3 из основных уравнений электромагнитной теории была получена геометрическая модель распространения света и было показано, что в некотором приближении изменение интенсивности в пучке света можно описать как функцию площади поперечного сечения трубки лучей. При суперпозиции двух или большего числа световых пучков распределение интенсивности в общем случае уже нельзя описывать таким простым способом. Так, если свет от одного источника разделить подходящим прибором на два пучка и затем наложить их друг на друга, то интенсивность в области суперпозиции пучков будет изменяться от точки к точке, достигая максимума, превышающего сумму интенсивностей пучков, и минимума, который может оказаться равным нулю. Это явление называется *интерференцией*. Ниже мы увидим, что при суперпозиции пучков строго монохроматического света интерференция возникает всегда. Однако свет от реальных физических источников никогда не бывает строго монохроматическим, и, как мы знаем из теории атомного строения, его амплитуда и фаза флуктуируют непрерывно и так быстро, что ни глаз, ни обычный физический детектор не могут уследить за их изменениями. Если два световых пучка происходят от одного источника, то возникающие в них флуктуации, вообще говоря, коррелированы, и о таких пучках говорят, что они полностью или частично *когерентны*, в зависимости от того, будет ли эта корреляция полной или частичной. В световых пучках от разных источников флуктуации совершенно независимы, и пучки, как говорят, взаимно *некогерентны*. При наложении таких пучков от независимых источников интерференция в обычных экспериментальных условиях не наблюдается, и полная интенсивность равна сумме интенсивностей отдельных пучков. Далее (см. гл. 10) мы увидим, что «степень корреляции» между флуктуациями в двух световых пучках определяет «четкость» интерференционных эффектов, возникающих при суперпозиции пучков, и наоборот, «степень корреляции» сама определяется этими эффектами.

Существуют два общих метода получения интерферирующих пучков из одного светового пучка, и они лежат в основе классификации устройств, применяемых в интерферометрии. В одном из них пучок делится, проходя сквозь близко расположенные друг к другу отверстия. Такой метод — *метод деления волнового фронта* — пригоден только при достаточно малых источниках. В другом способе пучок делится на одной или нескольких частично отражающих, частично пропускающих поверхностях. Этот метод — *метод деления амплитуды* — может применяться с протяженными источниками и обеспечивает большую интенсивность, чем первый метод. В любом случае удобно рассматривать отдельно явления, возникающие при суперпозиции двух пучков (*двухлучевая интерференция*), и явления, возникающие при суперпозиции большего их числа (*многолучевая интерференция*).

Интерференционные явления, исторически послужившие доказательством волновой теории света (см. «Историческое введение»), и в наши дни имеют важные практические применения, например в спектроскопии и метрологии. В настоящей главе мы коснемся главным образом идеализированного случая интерференции световых пучков от строго монохроматических источников. Эле-

ментарная теория монохроматических волн вполне достаточна для описания действия приборов, применяемых в подавляющем большинстве интерференционных исследований. В случае необходимости мы с помощью теоремы Фурье учтем, что реальные источники далеко не монохроматичны. Это обстоятельство всегда подразумевается в нашем представлении о протяженных источниках, которые мы считаем состоящими из большого числа взаимно некогерентных точечных источников. В настоящей главе мы всегда (когда это возможно) будем предполагать, что поведение отдельных пучков лучей подчиняется законам геометрической оптики и будем пренебрегать дифракционными явлениями, которые, как кратко указано в п. 3.1.4, возникают вблизи фокальных точек и границ тени. Эти явления будут подробно рассматриваться в гл. 8 для случая монохроматического света. Общий случай интерференции и дифракции полихроматического частично когерентного света будет рассмотрен в гл. 10.

§ 7.2. Интерференция двух монохроматических волн

Интенсивность света I определялась как усредненное по времени количество энергии, пересекающее единицу площади, перпендикулярной к направлению потока энергии, в единицу времени. Для плоской волны, согласно (1.4.8) и (1.4.9), имеем

$$I = v\langle w\rangle = \frac{c}{4\pi}\sqrt{\varepsilon/\mu}\,\langle \mathbf{E}^2\rangle = \frac{c}{4\pi}\sqrt{\frac{\mu}{\varepsilon}}\,\langle \mathbf{H}^2\rangle. \tag{1}$$

В гл. 3 мы видели, что эти соотношения справедливы, по крайней мере как приближения, и для волн более общего типа. Так как мы будем сравнивать интенсивности в одной и той же среде, можно считать величину $\langle \mathbf{E}^2\rangle$ мерой интенсивности. Большей частью мы будем иметь дело с монохроматическим полем, и поэтому представим электрический вектор $\mathbf{E}$ в виде

$$\mathbf{E}(\mathbf{r}, t) = \mathrm{Re}\,\{\mathbf{A}(\mathbf{r})\exp(-i\omega t)\} = \tfrac{1}{2}\,[\mathbf{A}(\mathbf{r})\exp(-i\omega t) + \mathbf{A}^*(\mathbf{r})\exp(i\omega t)]. \tag{2}$$

Здесь $\mathbf{A}$ — комплексный вектор с декартовыми компонентами

$$A_x = a_1(\mathbf{r})\exp[ig_1(\mathbf{r})],\quad A_y = a_2(\mathbf{r})\exp[ig_2(\mathbf{r})],\quad A_z = a_3(\mathbf{r})\exp[ig_3(\mathbf{r})], \tag{3}$$

где a_j и $g_j (j=1, 2, 3)$ — вещественные функции. Для однородной плоской волны амплитуды a_j постоянны, тогда как фазовые функции g_j имеют вид $g_j(\mathbf{r}) = \mathbf{k}\cdot\mathbf{r} - \delta_j$, где $\mathbf{k}$ — волновой вектор, а δ_j — фазовые постоянные, определяющие состояние поляризации. Из (2) имеем

$$\mathbf{E}^2 = \tfrac{1}{4}\,(\mathbf{A}^2\exp[-2i\omega t] + \mathbf{A}^{*2}\exp[2i\omega t] + 2\mathbf{A}\cdot\mathbf{A}^*). \tag{4}$$

Отсюда, усредняя по времени в интервале, большом по сравнению с периодом $T = 2\pi/\omega$, находим

$$\langle \mathbf{E}^2\rangle = \tfrac{1}{2}\mathbf{A}\cdot\mathbf{A}^* = \tfrac{1}{2}\,(|A_x|^2 + |A_y|^2 + |A_z|^2) = \tfrac{1}{2}\,(a_1^2 + a_2^2 + a_3^2). \tag{5}$$

Предположим теперь, что в некоторой точке P происходит суперпозиция двух монохроматических волн $\mathbf{E}_1$ и $\mathbf{E}_2$. Результирующее электрическое поле в P имеет вид

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 \tag{6}$$

и, следовательно,

$$\mathbf{E}^2 = \mathbf{E}_1^2 + \mathbf{E}_2^2 + 2\mathbf{E}_1\cdot\mathbf{E}_2. \tag{7}$$

Таким образом, полная интенсивность в точке P равна

$$I = I_1 + I_2 + J_{12}, \tag{8}$$

где

$$I_1 = \langle \mathbf{E}_1^2\rangle,\quad I_2 = \langle \mathbf{E}_2^2\rangle \tag{9а}$$

— интенсивности каждой из волн, а

$$J_{12} = 2\,\langle \mathbf{E}_1\cdot\mathbf{E}_2\rangle \tag{9б}$$

— *интерференционный член.* Пусть **A** и **B** — комплексные амплитуды волн, причем

$$A_x = a_1 e^{ig_1}, \quad \ldots, \quad B_x = b_1 e^{ih_1}, \quad \ldots \tag{10}$$

Вещественные фазы g_j и h_j обеих волн будут, вообще говоря, различны, так как волны приходят в P разными путями, но если условия эксперимента таковы, что между соответствующими компонентами возникла одна и та же *разность фаз* δ, то

$$g_1 - h_1 = g_2 - h_2 = g_3 - h_3 = \delta = \frac{2\pi}{\lambda_0} \Delta\mathcal{S}. \tag{11}$$

Здесь $\Delta\mathcal{S}$ — оптическая разность хода двух волн от общего источника до P, а λ_0 — длина волны в вакууме. Выражая произведение $\mathbf{E}_1 \cdot \mathbf{E}_2$ через **A** и **B**, найдем

$$\begin{aligned} \mathbf{E}_1 \cdot \mathbf{E}_2 &= {}^1/_4 (\mathbf{A} \exp(-i\omega t) + \mathbf{A}^* \exp(i\omega t)) (\mathbf{B} \exp(-i\omega t) + \mathbf{B}^* \exp(i\omega t)) = \\ &= {}^1/_4 (\mathbf{A} \cdot \mathbf{B} \exp(-2i\omega t) + \mathbf{A}^* \mathbf{B}^* \exp(2i\omega t) + \mathbf{A} \cdot \mathbf{B}^* + \mathbf{A}^* \cdot \mathbf{B}), \end{aligned} \tag{12}$$

и, следовательно,

$$\begin{aligned} J_{12} = 2 \langle \mathbf{E}_1 \cdot \mathbf{E}_2 \rangle &= {}^1/_2 (\mathbf{A} \cdot \mathbf{B}^* + \mathbf{A}^* \cdot \mathbf{B}) = a_1 b_1 \cos(g_1 - h_1) + \\ &+ a_2 b_2 \cos(g_2 - h_2) + a_3 b_3 \cos(g_3 - h_3) = (a_1 b_1 + a_2 b_2 + a_3 b_3) \cos\delta. \end{aligned} \tag{13}$$

Это выражение показывает зависимость интерференционного члена от амплитуд компонент и разности фаз обеих волн.

При выводе (13) мы не воспользовались электромагнитной теорией и, в частности, тем, что колебания поперечны. Как уже упоминалось в «Историческом введении», Френель и Араго показали, что два световых пучка, поляризованных под прямым углом друг к другу, не интерферируют и, следовательно, световые колебания должны быть поперечными. Это заключение легко вывести из (13). Предположим, что две волны распространяются в направлении z, и электрический вектор первой волны лежит в плоскости xz, а второй — в плоскости yz. Тогда $a_2 = 0$, $b_1 = 0$, и из (13) находим для интерференционного члена

$$J_{12} = a_3 b_3 \cos\delta.$$

Так как наблюдения Френеля и Араго показали отсутствие интерференции при таких условиях, мы должны заключить, что $a_3 = b_3 = 0$, т. е. электрические векторы обеих волн перпендикулярны направлению z. Следовательно, световые волны должны быть поперечными, что полностью согласуется с нашими прежними выводами из электромагнитной теории.

Рассмотрим распределение интенсивности в результате суперпозиции двух волн, распространяющихся в направлении оси z; пусть они линейно поляризованы и вектор **E** направлен по оси x. Тогда

$$a_2 = a_3 = b_2 = b_3 = 0,$$

и, используя (5), (9а) и (13), находим

$$I_1 = \frac{1}{2} a_1^2, \quad I_2 = \frac{1}{2} b_1^2, \quad J_{12} = a_1 b_1 \cos\delta = 2\sqrt{I_1 I_2} \cos\delta. \tag{14}$$

Тогда полная интенсивность, согласно (8), имеет вид

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\delta. \tag{15}$$

Очевидно, максимумы интенсивности равны

$$\left.\begin{aligned} I_{\text{макс}} &= I_1 + I_2 + 2\sqrt{I_1 I_2}, \\ \text{и появляются при} \quad |\delta| &= 0,\ 2\pi,\ 4\pi,\ \ldots, \end{aligned}\right\} \tag{16а}$$

а минимумы интенсивности равны

и появляются при
$$\left.\begin{array}{l} I_{мин} = I_1 + I_2 - 2\sqrt{I_1 I_2}, \\ |\delta| = \pi,\ 3\pi,\ \ldots \end{array}\right\} \qquad (16б)$$

В особом случае $I_1 = I_2$, соотношение (15) переходит в

$$I = 2I_1(1 + \cos\delta) = 4I_1 \cos^2\frac{\delta}{2}, \qquad (17)$$

и интенсивность изменяется от минимального значения $I_{мин} = 0$ до максимального значения $I_{макс} = 4I_1$ (рис. 7.1).

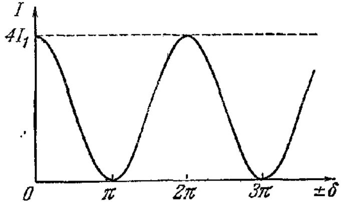


Рис. 7.1. Изменение интенсивности в зависимости от разности фаз при интерференции двух пучков равной интенсивности.

Эти формулы справедливы также и для естественного неполяризованного света, поскольку ниже (см. п. 10.8.9) будет показано, что пучок естественного света можно рассматривать как результат суперпозиции двух некогерентных пучков, линейно поляризованных под прямым углом друг к другу (например, по направлению x и y). В этом случае интерференцию между x- и y-компонентами следует рассматривать отдельно, и полная интенсивность получается сложением отдельных интенсивностей, а так как δ одинакова в каждом случае, то мы и находим приведенную выше формулу.

§ 7.3. Двухлучевая интерференция. Деление волнового фронта

7.3.1. Опыт Юнга. Первая экспериментальная установка для демонстрации интерференции света была осуществлена Юнгом. Свет от точечного монохроматического источника S падал на два небольших отверстия S_1 и S_2 в экране $\mathcal{A}$, расположенных рядом и находящихся на равных расстояниях от источника S (рис. 7.2). Эти отверстия действуют как вторичные монохроматические точечные и синфазные источники *), а световые пучки от них перекрываются позади экрана $\mathcal{A}$. Интерференция наблюдается в области перекрытия световых пучков.

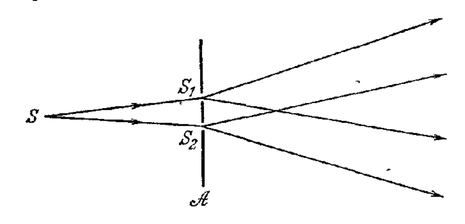


Рис. 7.2. Опыт Юнга.

Предположим, что интерференционная картина наблюдается в плоскости xOy, нормальной к перпендикуляру CO, восстановленному в середине отрезка, соединяющего точки S_1 и S_2, а ось x выберем параллельной S_1S_2 (рис. 7.3). Пусть d — расстояние между отверстиями, а a — расстояние между отрезком S_1S_2 и плоскостью наблюдения. Для точки P (x, y), лежащей в плоскости наблюдения, имеем

$$s_1 = S_1P = \sqrt{a^2 + y^2 + \left(x - \frac{d}{2}\right)^2}, \qquad (1а)$$

$$s_2 = S_2P = \sqrt{a^2 + y^2 + \left(x + \frac{d}{2}\right)^2}, \qquad (1б)$$

*) Такие источники создают направленное излучение, о чем подробно говорится в теории дифракции (см. гл. 8).

следовательно,

$$s_2^2 - s_1^2 = 2xd. \tag{2}$$

Разность геометрических путей света от S_2 и S_1 до P можно представить в виде

$$\Delta s = s_2 - s_1 = \frac{2xd}{s_2 + s_1}. \tag{3}$$

Вследствие малости длин волн видимого света интерференционная картина

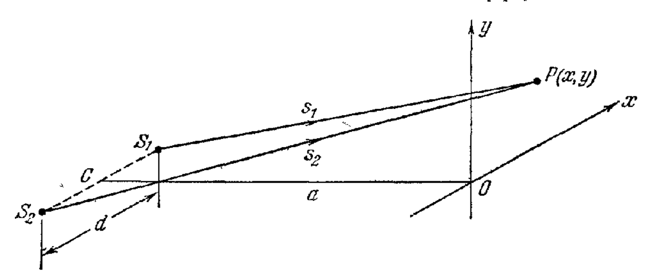


Рис. 7.3. Интерференция света от двух точечных источников.

будет наблюдаться, только если d значительно меньше a. Тогда при условии, что x и y также малы по сравнению с a, находим

$$s_2 + s_1 \sim 2a \tag{4}$$

и, отбрасывая члены второй и высших степеней величин d/a, x/a и y/a,

$$\Delta s = xd/a. \tag{5}$$

Если n — показатель преломления среды (предполагается, что она однородна), в которой происходит опыт, то оптическая разность хода от S_2 и S_1 до P равна

$$\Delta\mathcal{S} = n\Delta s = \frac{nxd}{a}, \tag{6}$$

а соответствующая разность фаз —

$$\delta = \frac{2\pi}{\lambda_0}\frac{nxd}{a}. \tag{7}$$

Так как угол S_1PS_2 очень мал, то допустимо считать, что волны от S_1 и S_2 движутся к P по одному и тому же направлению, и интенсивность можно рассчитать по формуле (7.2.15); согласно (7) и (7.2.16) максимумы интенсивности будут при

$$x = \frac{ma\lambda_0}{nd}, \quad |m| = 0,\ 1,\ 2,\ \ldots \tag{8а}$$

а минимумы интенсивности при

$$x = \frac{ma\lambda_0}{nd}, \quad |m| = \frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots \tag{8б}$$

Таким образом, интерференционная картина в непосредственной близости от O состоит из светлых и темных полос, называемых *интерференционными полосами* (рис. 7.4). Они находятся на равных расстояниях друг от друга и направлены под прямым углом к линии S_1S_2, соединяющей оба источника. Расстояние между соседними яркими полосами равно $a\lambda_0/nd$. В любой точке

интерференционной картины число m, определяемое соотношением

$$m=\frac{\delta}{2\pi}=\frac{\Delta\mathcal{S}}{\lambda_0}, \tag{9}$$

называется порядком интерференции в этой точке; следовательно, светлым полосам соответствуют целые порядки

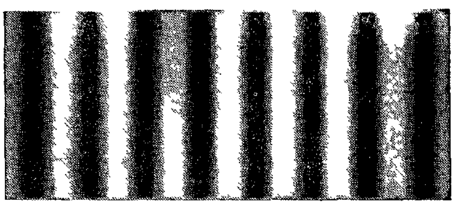


Рис 7 4 Интерференционные полосы в опыте Юнга

7.3.2. Зеркала Френеля и другие аналогичные устройства. Исторически опыты Юнга сыграли важную роль в становлении волновой теории света Они также дали метод (хотя и небольшой точности) измерения длины волны монохроматического света с помощью исключительно простой аппаратуры Для этого необходимо только измерить d, a и расстояния между полосами, которые в воздухе ($n\sim 1$) равны $a\lambda_0/d$ Однако в таком опыте свет от первичного источника S достигает области взаимодействия пучков не теми путями, которые описываются законами геометрической оптики, чтобы показать несущественность этого обстоятельства для осуществления интерференционных эффектов, впоследствии было предложено много других способов, позволяющих получать когерентные источники

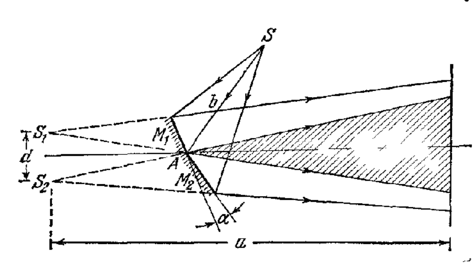


Рис 7 5 Зеркала Френеля

В качестве одного из примеров назовем устройство, известное под названием *зеркал Френеля* (рис 7 5) Свет от точечного источника S падает на два плоских зеркала M_1 и M , расположенные под небольшим углом друг к другу и, отражаясь от них, образует два мнимых изображения источника S_1 и S_2, которые действуют как когерентные источники Плоскость SS_1S_2, очевидно, нормальна к линии пересечения зеркал и пересекает ее в точке A. Если $SA=b$, то

$$S_1A=S_2A=b,$$

и, значит, перпендикуляр к середине отрезка S_1S_2 также проходит через A. Расстояние между S_1 и S_2 равно

$$d=2b\sin\alpha, \tag{10}$$

где α — угол между зеркалами

Рис 7 6 Зеркало Ллойда.

Еще проще устроено *зеркало Ллойда* (рис. 7.6). Точечный источник S_1 помещается на некотором расстоянии от плоского зеркала M очень близко к плоскости его поверхности, так что свет отражается зеркалом под углом, очень близким к скользящему Здесь когерентными источниками служат первичный источник S_1 и его мнимое изображение в зеркале S_2 При этом перпендикуляр к середине отрезка S_1S_2 лежит в плоскости зеркала.

Можно упомянуть еще два других устройства подобного рода. *Бипризма Френеля* (рис. 7.7) образуется двумя одинаковыми призмами с небольшим преломляющим углом, которые сложены основаниями и имеют параллельные преломляющие ребра. Пучок лучей от точечного источника S делится в результате преломления на два перекрывающихся пучка. Преломленные пучки не строго стигматичны, но вследствие малости преломляющего угла и малости угловой апертуры пучков мы можем пренебречь этой аберрацией и считать, что призмы образуют два мнимых изображения S_1 и S_2 источника S. *Билинза Бийе* (рис. 7.8) состоит из выпуклой линзы, разрезанной по диаметру на две части, немного раздвинутые в направлении, перпендикулярном к оптической оси; они образуют действительные изображения S_1 и S_2 точечного источника S.

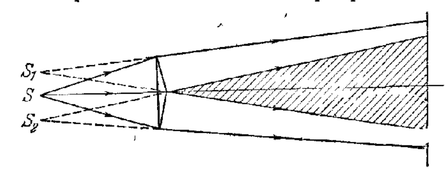


Рис. 7.7. Бипризма Френеля.

Во всех таких устройствах с первичным точечным источником интерференционные полосы наблюдаются в монохроматическом свете в любой плоскости области перекрытия расходящихся пучков от источников S_1 и S_2 (показано штриховкой на рис. 7.5—7.8). Про такие полосы говорят, что они *не локализованы.*

Обозначая, как и раньше, расстояния S_1P и S_2P через s_1 и s_2, найдем, что геометрическое место точек P, для которых разность фаз для волн, идущих от S_2 и S_1, постоянна, представляет собой поверхность, определяемую уравнением

$$s_2 - s_1 = \text{const}. \tag{11}$$

Следовательно, максимумы и минимумы результирующей интенсивности образуют семейство гиперболоидов с осью вращения S_1S_2, имеющих общие фокусы в S_1 и S_2. Интерференционные полосы в плоскости, нормальной к перпендикуляру CO, восстановленному к середине отрезка S_1S_2,— это сечения таких гиперболоидов, представляющие собой гиперболы. Однако вблизи O они, как мы видели в п. 7.3.1, близки к прямым эквидистантным линиям, направленным под прямым углом к S_1S_2. В плоскости, нормальной к S_1S_2, сечения гиперболоидов имеют вид концентрических окружностей, но такие полосы невозможно наблюдать с помощью описанных выше устройств. Однако их можно наблюдать в *опыте Меслина*, где половинки разрезанной линзы Бийе смещены вдоль оси, а не в поперечном направлении (рис. 7.9). В этом случае источники S_1 и S_2 расположены в разных точках оптической оси, и соответствующие световые пучки перекрываются в области между ними. Полосы могут наблюдаться в плоскостях, нормальных к S_1S_2, в виде концентрических окружностей с центром на оси. Так как область перекрытия пучков ограничена плоскостью, проходящей через оптическую ось, то видны только полуокружности.

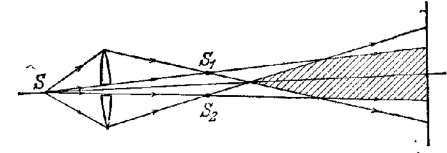


Рис. 7.8. Билинза Бийе.

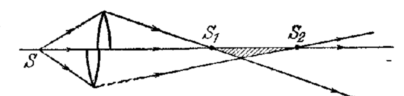


Рис. 7.9. Опыт Меслина.

7.3.3. Интерференционные полосы в квазимонохроматическом и белом свете. До сих пор мы предполагали, что первичный точечный источник испускает монохроматическое излучение. Теперь мы это ограничение снимем

и предположим, например, что зеркала Френеля освещены полихроматическим светом от точечного источника S (см. рис. 7.5). Как мы увидим далее (см. п. 7.5.8), такой свет можно представить как совокупность некогерентных монохроматических компонент, занимающих некоторый частотный диапазон. Каждая компонента образует свою интерференционную картину, аналогичную описанной выше, а полная интенсивность в любой точке равна сумме интенсивностей в таких монохроматических картинах. Предположим, что диапазон длин волн источника равен $\Delta\lambda_0$, а средняя длина волны $\bar{\lambda}_0$. Центральные максимумы всех монохроматических интерференционных картин, соответствующие равенству путей от S_1 и S_2, совпадают в точке O, но в любом другом месте появляющиеся картины смещены друг относительно друга, ибо их масштаб пропорционален длине волны. Максимумы m-го порядка займут в плоскости наблюдения участок Δx, равный, согласно (8),

$$\Delta x = \frac{|m|\,a}{nd}\,\Delta\lambda_0. \tag{12}$$

Рассмотрим сначала случай, когда диапазон длин волн $\Delta\lambda_0$ мал по сравнению со средней длиной волны $\bar{\lambda}_0$, т. е. когда

$$\frac{\Delta\lambda_0}{\bar{\lambda}_0} \ll 1. \tag{13}$$

Назовем свет, удовлетворяющий этому условию, *квазимонохроматическим светом* *).

Если для поля наблюдения

$$|m| \ll \frac{\bar{\lambda}_0}{\Delta\lambda_0} \tag{14}$$

или, согласно (9),

$$|\Delta\mathscr{S}| \ll \frac{\bar{\lambda}_0^2}{\Delta\lambda_0}, \tag{15}$$

то можно пренебречь величиной Δx по сравнению со средним расстоянием $a\bar{\lambda}_0/nd$ между соседними максимумами и считать, что для всех компонент интерференционные картины совпадают. Тогда в плоскости наблюдения появляются такие же полосы, как и в случае строго монохроматического света с длиной волны $\bar{\lambda}_0$.

Если свет квазимонохроматический, но условие (15) не выполняется, полосы будут менее отчетливы, чем в монохроматическом свете, и полная интенсивность будет зависеть от распределения интенсивностей среди монохроматических компонент.

Если свет не квазимонохроматичен, т. е. не выполняется условие (13), то наблюдаемая картина зависит также от спектральной чувствительности применяемого приемника излучения. В практически важном случае белого света и визуального наблюдения эффективный диапазон длин волн простирается приблизительно от 4000 до 7000 Å и $\Delta\lambda_0/\bar{\lambda}_0$ примерно равно 1/2. В этом случае на том месте, где при монохроматическом освещении должна была бы находиться полоса нулевого порядка, мы увидим центральную белую полосу; по обе стороны от нее располагаются окрашенные максимумы и минимумы, а за ними — пространство, которое кажется глазу равномерно освещенным белым светом. Однако это не обычный белый свет. Так, на расстоянии x от центральной

*) Так как $\lambda_0 = 2\pi c/\omega_0$, то (13) можно также записать в виде

$$\frac{\Delta\omega_0}{\bar{\omega}_0} \ll 1,$$

где $\Delta\omega_0$ — эффективный частотный диапазон, а $\bar{\omega}_0$ — средняя частота.

полосы находятся, согласно (8), максимумы интенсивности для

$$\lambda_0 = \frac{nd}{a}\frac{x}{m}, \quad |m| = 1,\ 2,\ 3,\ \ldots, \tag{16а}$$

и минимумы интенсивности для

$$\lambda_0 = \frac{nd}{a}\frac{x}{m}, \quad |m| = \frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots \tag{16б}$$

Следовательно, если такой свет падает на щель спектрографа, параллельную интерференционным полосам в монохроматическом свете (т. е. x остается постоянным для света, прошедшего сквозь щель), то спектр будет пересекаться светлыми и темными полосами, параллельными щели. Это один из примеров так называемого *канавчатого спектра*; расстояние между соседними светлыми полосами в нем равно $a/(nd|x|)$.

Из дальнейшего станет очевидным, что картина интерференционных полос в белом свете может оказаться полезной при интерферометрии, так как в некоторых особых случаях она позволяет обнаружить в монохроматическом свете полосу, соответствующую нулевой разности хода. Все, что мы говорили о центральной полосе (как о максимуме интенсивности), полученной с помощью зеркал Френеля, справедливо также для полос, создаваемых бипризмой Френеля, билинзой Бийе и устройством Юнга. Однако при использовании зеркала Ллоида полоса, лежащая в плоскости зеркала, соответствует минимуму интенсивности и в белом свете она кажется черной. Это объясняется тем, что при отражении от зеркала фаза волны независимо от ее длины меняется на π (см. п. 1.5 2). В опыте Меслина в центре картины также находится минимум интенсивности, так как (см. п. 8.8.4) после прохождения фокуса волна меняет фазу на π.

7.3.4. Источник в виде щели; видность полос. Все сказанное выше относилось к точечному первичному источнику. Однако все реальные источники имеют конечные размеры, и поэтому необходимо выяснить влияние размеров на интерференционную картину. Описание реальных физических источников требует привлечения теории атомного строения, выходящей за круг вопросов, рассматриваемых в настоящей книге, однако для наших целей можно пользоваться идеализированным представлением и считать, что источники состоят из большого числа точечных взаимно некогерентных элементов. Интенсивность в любой точке волнового поля равна тогда сумме интенсивностей от каждого точечного источника.

В устройствах, описанных выше (исключая опыт Меслина), интерференционные полосы перпендикулярны к плоскости, в которой находятся первичный точечный источник S и вторичные источники S_1 и S_2, и, следовательно, если S смещать перпендикулярно к этой плоскости, то полосы будут просто смещаться вдоль своих направлений. Таким образом, использование линейного источника (или на практике достаточно узкой щели), расположенного в этом направлении, не приведет к ухудшению четкости интерференционных полос, по крайней мере в тех случаях, когда их кривизна незначительна. Аналогично отверстия в опыте Юнга можно заменить узкими щелями, параллельными щели источника. Таким путем можно увеличить интенсивность интерференционной картины, хотя здесь и возникает дополнительная трудность, связанная с ориентацией источника.

Для получения более яркой картины необходимо увеличить ширину щели источника, но это приводит к тому, что полосы станут менее четкими. Рассмотрим в качестве примера зеркала Френеля. Если источник S сместить в положение S' под прямым углом к SA в плоскости SS_1S_2 (рис. 7.10), то и вторичные источники S_1 и S_2 переместятся в положение S_1' и S_2' и с точностью до членов второго порядка расстояние d между ними и, следовательно расстояние между полосами, не изменится; однако центральная полоса переместится из точки O

в точку O', так как она расположена в месте пересечения перпендикуляра к середине отрезка $S_1'S_2'$, проходящего через A, с экраном. Если $SS' = \zeta$ (считая направление, указанное на рис. 7.10, положительным), то $S_1S_1' = S_2S_2' = \zeta$ и

$$OO' = \frac{(a-b)\zeta}{b}. \tag{17}$$

Следовательно, если оптическая разность хода в P для света от источника S равна $\Delta\mathcal{S}$, то, согласно (6), оптическая разность хода для света от источника S' составит

$$\Delta\mathcal{S}' = \Delta\mathcal{S} - \frac{nd}{a}OO' = \Delta\mathcal{S} - D\zeta, \tag{18}$$

где

$$D = \frac{(a-b)\,nd}{ab}. \tag{19}$$

Соответствующая разность фаз записывается в виде

$$\delta(\zeta, \Delta\mathcal{S}) = \frac{2\pi}{\lambda_0}(\Delta\mathcal{S} - D\zeta). \tag{20}$$

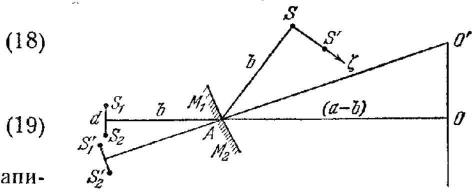


Рис. 7.10. Зеркала Френеля, освещаемые источником в виде щели.

Пусть источником служит щель шириной e с центром в S, и пусть число точечных источников, образующих щель, так велико, что ее можно считать непрерывным источником. Представим себе далее, что щель разделена на элементарные полоски, перпендикулярные к плоскости SS_1S_2. Если $i_1 d\zeta$ — интенсивность в точке P света, посылаемого элементарной полоской с помощью лишь одного зеркала, то интенсивность в этой точке света, посылаемого элементарной полоской S', равна, согласно (7.2.17),

$$i(\zeta, \Delta\mathcal{S})\,d\zeta = 2i_1(1 + \cos\delta)\,d\zeta,$$

а полная интенсивность в P — величине

$$I(e, \Delta\mathcal{S}) = 2i_1 \int_{-e/2}^{e/2} (1 + \cos\delta)\,d\zeta. \tag{21}$$

Подставляя (20) в (21) и вычисляя интеграл, получим

$$I(e, \Delta\mathcal{S}) = 2I_1\left\{1 + \left(\frac{\sin\frac{\pi De}{\lambda_0}}{\frac{\pi De}{\lambda_0}}\right)\cos\left(\frac{2\pi}{\lambda_0}\Delta\mathcal{S}\right)\right\}, \tag{22}$$

где $I_1 = i_1 e$.

Следуя Майкельсону, примем за меру четкости полос в точке P их «видность» $\mathcal{V}$, определяемую как

$$\mathcal{V} = \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}}, \tag{23}$$

где $I_{\text{макс}}$ и $I_{\text{мин}}$ — максимальная и минимальная интенсивности в непосредственной близости от P. Очевидно, $\mathcal{V}$ достигает максимальной величины, равной единице при $I_{\text{мин}} = 0$, как и в случае интерференции света от двух одинаковых монохроматических точечных источников, и уменьшается до нуля, когда $I_{\text{макс}} = I_{\text{мин}}$ и полосы исчезают. В нашем случае

$$I_{\text{макс}} = 2I_1\left(1 + \frac{\sin\frac{\pi De}{\lambda_0}}{\frac{\pi De}{\lambda_0}}\right), \quad I_{\text{мин}} = 2I_1\left(1 - \frac{\sin\frac{\pi De}{\lambda_0}}{\frac{\pi De}{\lambda_0}}\right), \tag{24}$$

и, значит,

$$\mathcal{V} = \frac{\left|\sin \frac{\pi De}{\lambda_0}\right|}{\frac{\pi De}{\lambda_0}}. \qquad (25)$$

Из (25) следует, что видность $\mathcal{V}$ зависит от ширины источника e (рис. 7.11). Мы видим, что видность превышает 0,9, если e меньше $\lambda_0/4D$ или, согласно (20), если разность фаз в точке P от элементов источника не больше $\pi/2$. Если с помощью этого несколько произвольного условия определить максимальную ширину щели, при которой видность полос еще остается удовлетворительной, то на основании (19) получим $e \leqslant \lambda_0 ab/4\ (a-b)nd$ или, используя (10),

$$e \leqslant \frac{\lambda_0 a}{8(a-b)\,n \sin\alpha}, \qquad (26)$$

где, как и раньше, α — угол между зеркалами.

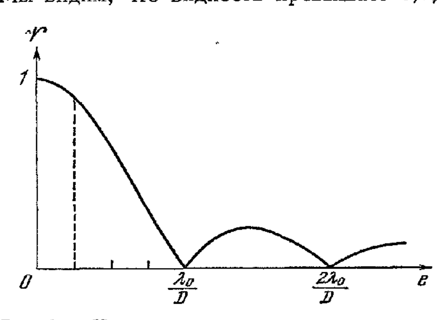


Рис. 7.11. Изменение видности полос при увеличении ширины источника в виде щели (зеркала Френеля).

Разумеется, допустимая величина e увеличивается с уменьшением $(a-b)$, т. е. по мере приближения плоскости наблюдения к линии пересечения зеркал. Например, в типичном случае $\alpha = 2'$, $a = 120$ *см*, $b = 100$ *см*, $\lambda_0 = 5500$ Å и $n = 1$ имеем $e \leqslant 0,7$ *мм*; расстояние между светлыми полосами, согласно (8) и (10), рассчитывается по формуле $a\lambda_0/2bn \sin\alpha$, и в нашем случае оно равно 0,57 *мм*.

Подобные же рассуждения относительно ширины источника применимы к бипризме Френеля, билинзе Бийе, к устройству Юнга. С зеркалом Ллойда положение иное, так как смещение источника S в направлении, перпендикулярном к плоскости зеркала, вызывает смещение его изображения в противоположном направлении. Следовательно, с источником в виде щели конечной ширины в элементарных картинах центральные минимумы в плоскости зеркала совпадают, но расстояние между минимумами различно, так что видность полос уменьшается с увеличением расстояния от плоскости зеркала.

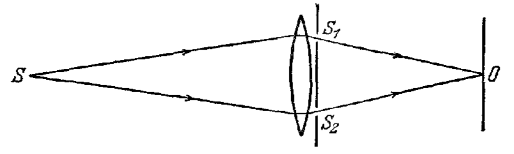


Рис. 7.12. Схема, иллюстрирующая применение линзы в устройстве Юнга.

7.3.5. Измерение оптической разности хода; интерферометр Рэлея. Воспользовавшись выводами теории дифракции, можно утверждать, что свет от вторичных источников S_1 и S_2 в опыте Юнга имеет наибольшую интенсивность в направлении геометрических лучей от первичного источника S. В опыте Юнга эти лучи за экраном расходятся, но с помощью линзы, поставленной перед отверстиями (рис. 7.12), их можно свести в точку O, сопряженную относительно линзы с S. Тогда интенсивность интерференционной картины вблизи O увеличивается, и можно наблюдать интерференционные полосы при отверстиях S_1 и S_2, находящихся значительно дальше друг от друга. Расстояние между соседними светлыми полосами по-прежнему равно $a\lambda_0/nd$, и если линза дает стигматическое изображение точки S, то, согласно принципу равенства оптического

пути, полоса нулевого порядка будет располагаться в O. Если же линза не дает стигматического изображения, полоса нулевого порядка сместится из O на величину, зависящую от оптической разности хода от S до O через оба отверстия. При оптической разности хода $\Delta\mathcal{S}$ смещение будет в Δm раз больше расстояния между соседними светлыми полосами, где

$$\Delta m = \frac{\Delta\mathcal{S}}{\lambda_0}. \tag{27}$$

Очевидно, что такое устройство можно использовать для количественного испытания качества линз, как это было сделано Майкельсоном [1]. Если одно из отверстий неподвижно относительно центра линзы, то, измеряя Δm при различных положениях другого отверстия, можно определить отклонение волнового фронта, идущего из S, от сферичности после прохождения линзы (волновая аберрация). Аналогично, если прозрачная пластинка толщиной l с показателем преломления n' помещена в пучок света, идущий от S_2, то оптическая длина пути $[SS_2O]$ увеличивается на $(n'-n)l$, и порядок интерференции в точке O изменится на величину

$$\Delta m = \frac{(n'-n)\,l}{\lambda_0}. \tag{28}$$

Измеряя Δm, l и λ_0, можно определить разность $(n'-n)$ между показателями преломления пластинки и окружающей среды. На этом основано устройство *интерферометра Рэлея* [2], применяемого для точных измерений показателей преломления газов. Схема современной модели этого прибора показана на рис. 7.13. Свет от щели S коллимируется линзой L_1 и затем падает на две другие щели S_1 и S_2, параллельные S. Параллельные пучки света от S_1 и S_2 проходят через разные газовые кюветы T_1 и T_2 и собираются линзой L_2, в фокальной плоскости которой образуются интерференционные полосы, параллельные щелям *). Помещение газовых кювет в пучки света заставляет значительно увеличить расстояние между щелями S_1 и S_2, вследствие чего интерференционные полосы располагаются тесно, и для их наблюдения требуется большое увеличение. Ширина щели S также не может быть большой, и, следовательно, яркость картины невелика. Так как увеличение требуется только в направлении, перпендикулярном к полосам, то для этой цели хорошо подходит цилиндрический окуляр в виде тонкой стеклянной палочки с длинной осью, параллельной полосам. Картина, рассматриваемая таким образом, значительно ярче, чем при использовании сферического окуляра. Применение цилиндрического окуляра имеет еще и другое важное преимущество, позволяя получить вторую фиксированную систему полос с таким же расстоянием между полосами, как и у главной, но образованную светом от источников S_1 и S_2, прошедшим ниже газовых кювет. Вторая система полос может служить шкалой для отсчета. С помощью стеклянной пластинки G эту шкалу смещают по вертикали так, чтобы ее верхний край соприкасался с нижним краем главной системы. Резкая линия раздела между ними — это край пластинки G, наблюдаемый через линзу

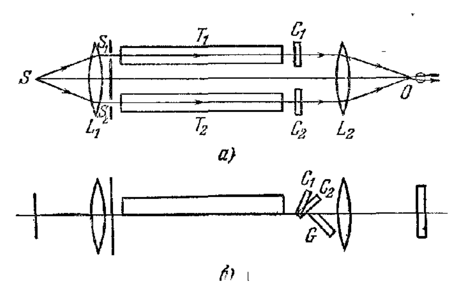


Рис. 7.13. Схема интерферометра Рэлея. *а* — горизонтальное сечение, *б* — вертикальное сечение.

*) Точное распределение интенсивности в фокальной плоскости линзы L_2 можно найти из теории дифракции Фраунгофера (см. § 8.5).

L_2. Следовательно, определение смещения главной системы полос, обусловленного изменением оптических путей в кюветах T_1 и T_2, целиком зависит от остроты зрения глаза, которая, вообще говоря, велика, и таким способом можно обнаружить смещения, примерно равные 1/40 порядка. Случайные смещения в оптической системе также становятся менее существенными, так как сказываются одновременно на обеих системах полос.

На практике удобнее компенсировать оптическую разность хода, а не считать полосы. Это делается следующим образом: свет, выходящий из газовых кювет, проходит через тонкие стеклянные пластинки, одна из которых C_1 неподвижна, а другая C_2 может вращаться вокруг горизонтальной оси, что позволяет плавно изменять оптическую длину пути света, выходящего из S_2. Такой компенсатор калибруется в монохроматическом свете для того, чтобы определить величину поворота пластинки, соответствующую смещению на один порядок в главной системе полос. В этом случае система полос служит нуль-индикатором равенства оптических путей $[SS_1O]$ и $[SS_2O]$. Обычно работа с прибором происходит следующим образом: газовые кюветы откачивают, и в белом свете с помощью компенсатора примерно совмещают полосы главной системы и шкалы; затем добиваются точного совпадения нулевых порядков в монохроматическом свете, после чего одну из кювет заполняют исследуемым газом и снова, сперва в белом свете, а потом в монохроматическом совмещают, используя компенсатор, нулевые порядки. Разница между двумя установками компенсатора позволяет определить по его калибровке смещение порядка Δm в главной системе полос, вызванное присутствием газа в кювете. Показатель преломления этого газа n' находят из (28), а именно:

$$(n'-1)=\frac{\lambda_0}{l}\Delta m, \tag{29}$$

где l — длина газовой кюветы. При обычных значениях $l=100$ *см*, $\lambda_0=5500$ Å и точности установки в 1/40 порядка можно обнаружить изменение $(n'-1)$ *) около 10^{-8}.

Оптические пути от S_1 и S_2 до места наблюдения интерференционной картины проходят среды с различной дисперсией; поэтому, в отличие от простого случая, рассмотренного в п. 7.3.3, нулевые порядки в свете разных длин волн, вообще говоря, не совпадают, и в белом свете отсутствует совершенно белая полоса. У наименее окрашенной полосы $\partial m/\partial\lambda_0=0$ для некоторой средней длины волны $\lambda=\bar{\lambda}_0$ (в видимой области спектра), которая зависит от цветовой чувствительности глаза. По аналогии с терминологией, принятой при описании линз, эта полоса называется *ахроматической*. Если компенсатор вводит оптическую разность хода $\Delta\mathcal{S}$, то порядок интерференции в точке O равен

$$m=[(n'-1]\,l+\Delta\mathcal{S}]\,\frac{1}{\lambda_0},$$

так что

$$\frac{\partial m}{\partial\lambda_0}=-\left(n'-1-\lambda_0\frac{\partial n'}{\partial\lambda_0}\right)\frac{1}{\lambda_0^2}-\left(\Delta\mathcal{S}-\lambda_0\frac{\partial}{\partial\lambda_0}(\Delta\mathcal{S})\right)\frac{1}{\lambda_0^2}.$$

Поэтому в точке O ахроматическая полоса будет тогда, когда

$$\left(\Delta\mathcal{S}-\lambda_0\frac{\partial}{\partial\lambda_0}(\Delta\mathcal{S})\right)_{\bar{\lambda}_0}=-l\left(n'-1-\lambda_0\frac{\partial n'}{\partial\lambda_0}\right)_{\bar{\lambda}_0}.$$

При такой установке компенсатора нулевой порядок картины в монохроматическом свете может не попасть в точку O, так как для их совпадения требуется, чтобы

$$(\Delta\mathcal{S})_{\lambda_0}=-(n'-1)_{\lambda_0}.$$

Это несовпадение может оказаться достаточно большим, чтобы затруднить идентификацию полосы нулевого порядка в монохроматическом свете, и поэтому приходится прибегать к предварительным измерениям при малом давлении или с короткой кюветой.

Заметим также, что ахроматическая полоса хорошо распознается, только если в тех точках картины, где $(\partial m/\partial\lambda_0)_{\bar{\lambda}_0}=0$, область значений m для длин волн видимого спектра достаточно мала. При наблюдении в белом свете пути интерферирующих волн в средах с одинаковой дисперсией должны быть по возможности равными.

*) Величина $(n'-1)$ для обычных газов порядка 10^{-4} (см. табл. 1.1).

Бо́льшую чувствительность в принципе можно получить, увеличивая l, но этому препятствуют трудности контроля температуры. По той же причине в модели прибора, предназначенного для измерения разности показателей преломления жидкостей, применяются только короткие кюветы. Кроме того, разность хода, которую можно скомпенсировать, ограничена, и поэтому при большой разнице показателей преломления в кюветах длина их должна быть пропорционально уменьшена.

7.3.6. Измерение угловых размеров источников; звездный интерферометр Майкельсона. Мы уже видели (п. 7.3.4), что в опыте Юнга четкость полос зависит от размеров источника в направлении, соединяющем отверстия S_1 и S_2. На этом эффекте основан метод измерения угловых размеров малых источников.

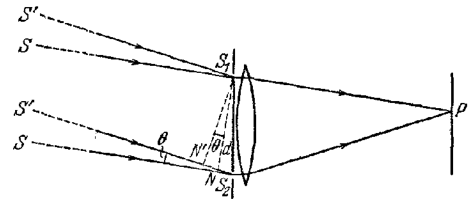


Рис. 7.14. Объектив телескопа, закрытый диафрагмой с двумя отверстиями и освещенный светом от двух удаленных точечных источников.

Предположим, что в телескоп, объектив которого закрыт диафрагмой с двумя небольшими отверстиями S_1 и S_2 с расстоянием d между ними, рассматриваются два удаленных квазимонохроматических точечных источника S и S', испускающих свет с эффективной длиной волны λ_0, причем угловое расстояние между ними равно θ (рис. 7.14). Каждый из источников S и S' даст в фокальной плоскости объектива интерференционную картину с одинаковым расстоянием между полосами. Если источники S и S' некогерентны, то общая картина образуется суммированием в каждой точке интенсивностей каждой картины. Пусть N — основание перпендикуляра, опущенного из S_1 на SS_2; тогда S_1N лежит в плоскости волнового фронта, идущего от источника S, так что в точке P в фокальной плоскости оптическая разность хода для света от S равна

$$\Delta\mathcal{S} = [SS_2P] - [SS_1P] = [NS_2] + [S_2P] - [S_1P]. \tag{30}$$

Аналогично, если N'— основание перпендикуляра, опущенного из S_1 на $S'S_2$, то оптическая разность хода в точке P для света от S' равна

$$\Delta\mathcal{S}' = [N'S_2] + [S_2P] - [S_1P]. \tag{31}$$

Следовательно, интерференционные картины от S' и S смещены одна относительно другой на Δm порядков, где

$$\Delta m = \frac{|\Delta\mathcal{S}' - \Delta\mathcal{S}|}{\lambda_0} = \frac{|[N'S_2] - [NS_2]|}{\lambda_0} \sim \frac{\theta d}{\lambda_0} \tag{32}$$

при $n \sim 1$ и малом θ. Когда

$$\Delta m = 0,\ 1,\ 2,\ \ldots, \tag{33a}$$

максимумы интенсивностей в интерференционной картине, создаваемой источниками S и S', совпадают, и полосы в общей картине будут наиболее четкими. В другом случае, когда

$$\Delta m = {}^1/_2,\ {}^3/_2,\ {}^5/_2,\ \ldots, \tag{33б}$$

максимумы интенсивностей в картине, получающейся от S, совпадают с минимумами интенсивности в картине от S'; тогда полосы в общей картине становятся наименее отчетливыми и исчезают при равенстве обеих интенсивностей (от S и S'). Таким образом, в изменении четкости полос наблюдается периодичность при увеличении расстояния между отверстиями от нуля. В частности, первый минимум отчетливости полос будет при

$$d = \lambda_0/(2\theta), \tag{34}$$

и, если он наблюдается, то отсюда можно определить θ, зная d и λ_0.

Рассмотрим теперь более общий случай протяженного квазимонохроматического первичного источника с центром в S. Как и прежде, предположим, что такой источник состоит из взаимно некогерентных точечных источников. Можно представить себе, что он состоит из элементарных полосок, расположенных, перпендикулярно к линии, соединяющей отверстия S_1 и S_2. Найдем полную интенсивность в точке P фокальной плоскости объектива, суммируя вклады интенсивности, обусловленные каждой такой полоской. Обозначая через $\Delta\mathscr{S}$ оптическую разность хода в P для света, идущего от элемента S источника, находим, что для данного d положение P определяется величиной $\Delta\mathscr{S}$. Оптическая разность хода для света, идущего от элемента, расположенного под углом α к S, равна, согласно (32), величине $\Delta\mathscr{S}+\alpha d$, а соответствующая разность фаз — величине

$$\delta(\alpha,\ \Delta\mathscr{S})=\frac{2\pi}{\lambda_0}(\Delta\mathscr{S}+\alpha d). \tag{35}$$

Допустим снова, что протяженный источник образован таким большим числом точечных источников, что мы можем считать его непрерывным. Тогда, согласно (7.2.15), полная интенсивность в точке P равна

$$I(d,\ \Delta\mathscr{S})=\int i_1\,d\alpha+\int i_2\,d\alpha+2\int\sqrt{i_1 i_2}\cos\delta\,d\alpha, \tag{36}$$

где $i_1(\alpha,\ \Delta\mathscr{S})\,d\alpha$ и $i_2(\alpha,\ \Delta\mathscr{S})\,d\alpha$ — интенсивности в точке P от одной элементарной плоскости, когда свет приходит в эту точку только через одно отверстие. Интегрирование проводится по всем значениям α, соответствующим всем элементам источника.

В п. 7.3.5 уже отмечалось, что вследствие дифракции вторичные источники, образованные отверстиями S_1 и S_2, имеют свойства направленности, однако если этим можно пренебречь *) в области значений α, входящих в (36), то мы вправе написать

$$i_1(\alpha,\ \Delta\mathscr{S})=i_2(\alpha,\ \Delta\mathscr{S})=i(\alpha)\,f(\Delta\mathscr{S}), \tag{37}$$

где $i(\alpha)$ пропорционально интенсивности соответствующих полосок источника, а $f(\Delta\mathscr{S})$ характеризует направленность излучения S_1 (или S_2). Используя (37) и (35), получим из соотношения (36)

$$I(d,\ \Delta\mathscr{S})=2f(\Delta\mathscr{S})\int i(\alpha)\{1+\cos\delta\}\,d\alpha=$$
$$=f(\Delta\mathscr{S})\left\{P+C(d)\cos\left(\frac{2\pi}{\lambda_0}\Delta\mathscr{S}\right)-S(d)\sin\left(\frac{2\pi}{\lambda_0}\Delta\mathscr{S}\right)\right\}, \tag{38}$$

где

$$\left.\begin{aligned}P&=2\int i(\alpha)\,d\alpha,\\ C(d)=2\int i(\alpha)\cos\left(\frac{2\pi}{\lambda_0}\alpha d\right)d\alpha,&\quad S(d)=2\int i(\alpha)\sin\left(\frac{2\pi}{\lambda_0}\alpha d\right)d\alpha.\end{aligned}\right\} \tag{39}$$

Если допустить еще, что изменение величины $f(\Delta\mathscr{S})$ мало по сравнению с изменением $\cos(2\pi\Delta\mathscr{S}/\lambda_0)$ и $\sin(2\pi\Delta\mathscr{S}/\lambda_0)$, то положения максимумов и минимумов I будут определяться соотношением *)

$$\frac{\partial I}{\partial(\Delta\mathscr{S})}=0=-\frac{2\pi}{\lambda_0}\left\{C\sin\left(\frac{2\pi}{\lambda_0}\Delta\mathscr{S}\right)+S\cos\left(\frac{2\pi}{\lambda_0}\Delta\mathscr{S}\right)\right\},$$

т. е.

$$\operatorname{tg}\left(\frac{2\pi}{\lambda_0}\Delta\mathscr{S}\right)=-\frac{S}{C}. \tag{40}$$

---

*) Это приближение оправдано, если отверстия достаточно узки, что будет показано в § 8.5.

Из (38) и (40) находим для экстремальных значений I

$$I_{\text{экстр}} = P \pm \left|\sqrt{C^2+S^2}\right|, \tag{41}$$

и, следовательно, видность полос, определяемая (23), равна

$$\mathcal{V}^{\mathcal{P}}(d) = \frac{\left|\sqrt{C^2+S^2}\right|}{P}. \tag{42}$$

Вид функции $\mathcal{V}^{\mathcal{P}}(d)$ для нескольких различных зависимостей $i(\alpha)$ показан на рис. 7.15. Рис. 7.15, *а* соответствует случаю двух точечных источников, которые мы кратко рассмотрели в начале этого раздела; рис. 7.15, *б* — однородному прямоугольному источнику со сторонами, параллельными линии, соединяющей отверстия S_1 и S_2; очевидно, в этом случае функция $\mathcal{V}^{\mathcal{P}}(d)$ совпадает

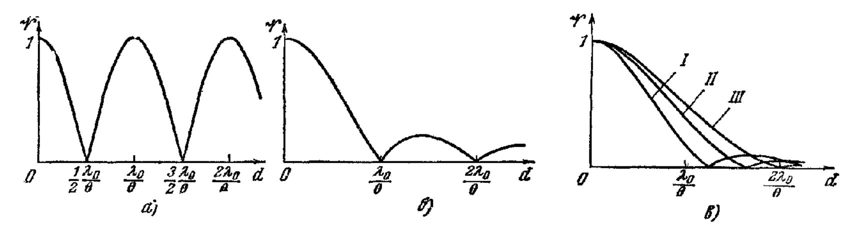


Рис. 7.15. Изменение видности полос в зависимости от расстояния между отверстиями (см. рис. 7.14).

а — два одинаковых точечных источника с угловым расстоянием θ, *б* - однородный прямоугольный источник угловой ширины θ со сторонами, параллельными линии, соединяющей отверстия *в* — источник в виде круглого диска с угловым диаметром $\theta = 2\beta_0$ и распределением интенсивности $I(\beta) \propto (\beta_0^2 - \beta^2)^v$, где β — угловой радиус, отсчитываемый от центра [3]

с функцией видности, показанной на рис. 7.11. На рис. 7.15, *в* источником служит круглый диск с радиальной симметрией. Кривая *I* соответствует постоянной интенсивности, кривые *II* и *III* — разным степеням ее убывания от центра к краю.

Можно рассмотреть и обратную задачу; если известны положение полос и их видность в зависимости от расстояния между отверстиями d, то из (40) и (42) определяются функции C и S, за исключением постоянного множителя пропорциональности P и его знака. Последний обычно находят из физических соображений. В таком случае распределение интенсивности $i(\alpha)$ по источнику получается из (39) с помощью обратной теоремы Фурье. Подобные измерения, хотя они принципиально и возможны, довольно трудны. Однако если заранее известно, что источник имеет одну из форм, соответствующих рис. 7.15, то его угловые размеры можно найти, просто определяя наименьшее значение d, при котором видность полос минимальная. Это условие осуществляется, когда

$$d = A\lambda_0/\theta, \tag{43}$$

где, как мы видели в (34), $A = 0{,}5$ для двух точечных источников с угловым расстоянием θ между ними; $A = 1{,}22$ для однородного круглого источника в форме диска с угловым диаметром θ и $A > 1{,}22$, если у краев диск темнее, чем в центре.

Описанный метод был предложен впервые Физо [14] и позднее Майкельсоном [5] для определения угловых размеров астрономических объектов, которые слишком малы или слишком далеки, чтобы это можно было сделать недиафрагмированным телескопом (см. п. 8.6.2). Такие объекты испускают белый свет, и поскольку его интенсивность очень мала, наблюдения должны проводиться

в белом свете *). Следовательно, необходимо предположить, что в (43) λ_0 — эффективная длина волны, зависящая от распределения интенсивности света по частотам и от цветовой чувствительности глаза. С такими ограничениями этим методом успешно пользовались для измерения угловых диаметров спутников планет [6] и угловых расстояний между компонентами двойных звезд, диаметры которых малы по сравнению с расстоянием между ними [7]. Однако попытки применить его для определения угловых диаметров одиночных звезд практически не удались, так как из-за малости этих диаметров полосы оставались четкими даже при наибольших расстояниях между отверстиями, еще допустимых с существовавшими тогда телескопами. Майкельсон [8] преодолел возникшие трудности, построив свой *звездный интерферометр* (рис. 7.16). Отверстия S_1 и S_2, диафрагмирующие объектив телескопа, неподвижны, и свет достигает их, отразившись от симметричной системы зеркал M_1, M_2, M_3 и M_4, установленных на жесткой ферме перед телескопом. Внутренние зеркала M_3 и M_4 неподвижны, а внешние M_1 и M_2 могут симметрично смещаться в направлении линии, соединяющей S_1 и S_2. Если оптические пути $[M_1M_3S_1]$ и $[M_2M_4S_2]$ равны, то оптическая разность хода для света от удаленного точечного источника одинакова как в S_1 и S_2, так и на M_1 и M_2, и, следовательно, внешние зеркала играют роль подвижных отверстий в методе Физо. Таким образом, наименьший угловой диаметр, который можно измерить подобным устройством, определяется не диаметром объектива телескопа, а максимальным расстоянием между внешними зеркалами. Другое достоинство данной системы заключается в том, что расстояние между полосами, которое зависит от расстояния между S_1 и S_2, остается постоянным при изменении расстояния между подвижными зеркалами. Такой интерферометр был смонтирован на большом отражательном телескопе (диаметр 2,5 *м*) обсерватории Маунт Вильсон, выбранном только из-за прочности своей механической конструкции. При расстоянии между отверстиями S_1 и S_2 в 114 *см* расстояние между полосами в фокальной плоскости равнялось примерно 0,02 *мм*.

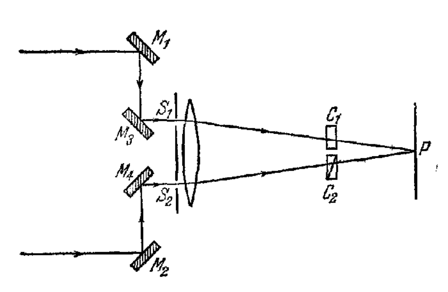


Рис. 7.16. Звездный интерферометр Майкельсона.

Максимальное расстояние между внешними зеркалами составляло 6,1 *м*, наименьший измеримый угловой диаметр (при $\lambda_0 = 5500$ Å) — около 0,02″. Вследствие неизбежных механических дефектов понадобилось еще два вспомогательных устройства, обеспечивающих правильную установку внешних зеркал во всех положениях. Чтобы свести геометрические световые пучки от S_1 и S_2 в фокус, применялась плоскопараллельная стеклянная пластинка C_1, которую можно было наклонять в любую сторону. Другая плоскопараллельная стеклянная пластинка C_2 переменной толщины использовалась для компенсации неравенства оптических путей $[M_1M_3S_1]$ и $[M_2M_4S_2]$. Эта компенсация необходима, ибо полосы в белом свете видимы только вблизи нулевого порядка. Ее контролировали, наблюдая канавчатый спектр в небольшой спектроскоп.

Первой звездой, у которой удалось измерить угловой диаметр, была Бетельгейзе (α Ориона). Полученная величина равнялась 0,047″. По расстоянию

*) Общая проблема наблюдения интерференционных полос от протяженного источника с бесконечным набором длин волн изящно разобрана в теории частично когерентного света (см. гл. 10). В рамках этой теории действие звездного интерферометра Майкельсона кратко рассматривается в § 10.4.

между этой звездой и Солнцем (определенному тригонометрически) был найден ее линейный диаметр, оказавшийся равным $4,1 \cdot 10^8$ *км*, что почти в 300 раз больше диаметра Солнца ($1,4 \cdot 10^6$ *км*) и превышает диаметр земной орбиты ($3 \cdot 10^8$ *км*). Таким способом были измерены диаметры только нескольких звезд. Все они, подобно Бетельгейзе,— гиганты с линейными диаметрами во много раз большими, чем у Солнца. Небольшое число измерений звездных диаметров частично объясняется трудностью наблюдений, связанной с турбулентным возмущением атмосферы, хотя Майкельсон и Пиз показали, что при наблюдении с интерферометром эти вредные влияния сказываются значительно меньше, чем при наблюдении с обычным телескопом большого диаметра. Изменение показателя преломления воздуха перед небольшими отверстиями интерферометра смещает всю интерференционную картину в целом и полосы остаются различимыми, если только это смещение происходит медленно. При наблюдении в таких же условиях в телескоп с недиафрагмированным объективом изображение звезды оказывается сильно искаженным. Однако дело не только в трудности наблюдений; расстояние между внешними зеркалами в 6 *м* совершенно недостаточно, так как подавляющее большинство звезд мало отличается по своему диаметру от Солнца. На расстоянии же ближайшей звезды солнечный диск был бы виден под углом лишь 0,007″ и для наблюдения первого исчезновения полос расстояние между зеркалами должно было бы равняться примерно 20 *м*. Постройка такого большого интерферометра трудна вследствие высоких требований, предъявляемых к жесткости механических соединений между зеркалами и окуляром.

Инструмент, подобный звездному интерферометру Майкельсона, используется в радиоастрономии для определения угловых размеров небесных радиоисточников [9]. Он состоит из двух разнесенных антенн, сигналы с которых подаются на общий детектор системы. Здесь технические трудности также возрастают по мере увеличения расстояния между антеннами вследствие внесения непостоянной разности фаз на пути от антенн к детектору. Браун и Твисс [10] создали другой тип радиоинтерферометра, свободный от этого недостатка. В предложенном ими устройстве сигналы, принятые антеннами, детектируются независимо, а угловые размеры источника находят из измерений корреляции в флуктуации интенсивности сигналов как функции расстояния между антеннами. Они также показали [11], что аналогичное устройство может работать и в видимом свете. Свет от звезды собирается двумя вогнутыми зеркалами и фокусируется на два фотоэлемента: корреляция флуктуаций фототоков измеряется в зависимости от расстояния между зеркалами. В таком устройстве (называемом *интерферометром интенсивностей*) большое расстояние между зеркалами не вызывает осложнений, и становится возможным измерение звезд значительно меньших диаметров. Принципы данного метода легче понять в рамках теории частичной когерентности (см. § 10.4).

§ 7.4. Стоячие волны

В устройствах, о которых говорилось до сих пор, две интерферирующие волны распространялись вблизи точки наблюдения почти в одном направлении. Теперь мы рассмотрим интерференцию двух волн, распространяющихся в противоположных направлениях, например интерференцию падающей и отраженной плоских монохроматических волн света при его падении на хорошо отражающую плоскую поверхность. Предположим, что такой поверхностью служит плоскость $z = 0$ с положительной осью z, направленной в сторону среды, в которой распространяется падающая волна, и обозначим через n_1 и n_2 показатели преломления обеих сред (рис. 7.17).

Пусть θ_i— угол падения, а xz — плоскость падения. Пусть далее $\mathbf{E}^{(i)}$ — электрический вектор падающей волны, а $A_\parallel$ и $A_\perp$, как и в п. 1.5.2 — его

параллельная и перпендикулярная компоненты относительно плоскости падения. Декартовы компоненты вектора $\mathbf{E}^{(i)}$ можно получить из (1.5.11), заменяя z на $-z$, т. е. используя, как обычно, только вещественные части,

$$\left.\begin{aligned} E_x^{(i)} &= -A_{\parallel}\cos\theta_i \exp(-i\tau_i),\\ E_y^{(i)} &= A_{\perp}\exp(-i\tau_i),\\ E_z^{(i)} &= -A_{\parallel}\sin\theta_i \exp(-i\tau_i), \end{aligned}\right\} \tag{1}$$

где

$$\tau_i = \omega\left(t - \frac{x\sin\theta_i - z\cos\theta_i}{v_1}\right), \tag{2}$$

а v_1 — скорость распространения в первой среде. Декартовы компоненты вектора $\mathbf{E}^{(r)}$ отраженной волны получаются из аналогичных выражений (см. уравнение (1.5.16), которые, если воспользоваться законом отражения (1.5.7), принимают вид

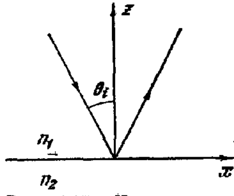


Рис. 7.17. К отражению плоской волны от плоской поверхности.

$$E_x^{(r)} = R_{\parallel}\cos\theta_i \exp(-i\tau_r),\qquad E_y^{(r)} = R_{\perp}\exp(-i\tau_r),$$
$$E_z^{(r)} = -R_{\parallel}\sin\theta_i \exp(i\tau_r), \tag{3}$$

где

$$\tau_r = \omega\left(t - \frac{x\sin\theta_i + z\cos\theta_i}{v_i}\right). \tag{4}$$

Амплитуды $R_{\parallel}$ и $R_{\perp}$ отраженной волны связаны с амплитудами $A_{\parallel}$ и $A_{\perp}$ падающей волны формулами Френеля (1.5.21), т. е.

$$R_{\parallel} = \frac{n_2\cos\theta_i - n_1\cos\theta_t}{n_2\cos\theta_i + n_1\cos\theta_t}A_{\parallel},\qquad R_{\perp} = \frac{n_1\cos\theta_i - n_2\cos\theta_t}{n_1\cos\theta_t + n_2\cos\theta_t}A_{\perp}. \tag{5}$$

Для упрощения предположим, что n_2/n_1 так велико, что отражательную способность можно принять за единицу; тогда в пределе $n_2/n_1 \to \infty$ последнее соотношение запишется в виде *)

$$R_{\parallel} = A_{\parallel},\qquad R_{\perp} = -A_{\perp}. \tag{6}$$

Подставляя (6) в (3) и добавляя выражение для падающего и отраженного полей, получим полное поле. Следовательно, x-компонента электрического вектора полного поля равна

$$E_x = E_x^{(i)} + E_x^{(r)} = 2A_{\parallel}\cos\theta_i\left\{\sin\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right) - \frac{\pi}{2}\right]\right\}. \tag{7а}$$

Аналогично

$$E_y = -2A_{\perp}\left\{\sin\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right) - \frac{\pi}{2}\right]\right\} \tag{7б}$$

и

$$E_z = -2A_{\parallel}\sin\theta_i\left\{\cos\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right)\right]\right\}. \tag{7в}$$

Точно таким же образом мы получим из уравнений (1.5.13) и (1.5.16) следующие выражения для компонент магнитного вектора полного поля:

$$H_x = -2A_{\perp}n_1\cos\theta_i\left\{\cos\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right)\right]\right\}, \tag{8а}$$

$$H_y = -2A_{\parallel}n_1\left\{\cos\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right)\right]\right\}, \tag{8б}$$

$$H_z = 2A_{\perp}n_1\left\{\sin\theta_i\sin\left(\frac{\omega z\cos\theta_i}{v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t - \frac{x\sin\theta_i}{v_1}\right) - \frac{\pi}{2}\right]\right\}, \tag{8в}$$

*) Мы исключили случай скользящего падения ($\cos\theta_i \to 0$), для которого отражательная способность приближается к единице, даже когда n_2/n_1 невелико; в этом случае (5) дает $R_{\parallel} = -A_{\parallel}$, $R_{\perp} = -A_{\perp}$, что соответствует случаю зеркала Ллойда (см. п. 7.3.2).

где использовано соотношение Максвелла $n_1=\sqrt{\varepsilon_1}$. Каждое из выражений (7) и (8) представляет волну, распространяющуюся в направлении x со скоростью $v_1/\sin\theta_t$. Амплитуда этой волны непостоянна и периодически изменяется в направлении z с периодом $2\pi v_1/(\omega\cos\theta_t)=\lambda_0/(n_1\cos\theta_t)$, где λ_0—длина волны в вакууме.

Случай нормального падения ($\theta_t=0$) особенно интересен. Если A_x и A_y— амплитуды компонент электрического вектора, то, согласно (1), можно написать $A_\parallel=-A_x$, $A_\perp=A_y$; тогда из (7) и (8) находим

$$\left.\begin{aligned} E_x&=-2A_x\left\{\sin\left(\frac{\omega z}{v_1}\right)\right\}\exp\left[-i\left(\omega t-\frac{\pi}{2}\right)\right],\\ E_y&=-2A_y\left\{\sin\left(\frac{\omega z}{v_1}\right)\right\}\exp\left[-i\left(\omega t-\frac{\pi}{2}\right)\right],\quad E_z=0\end{aligned}\right\}\tag{9}$$

и

$$\left.\begin{aligned} H_x&=-2A_yn_1\left\{\cos\left(\frac{\omega z}{v_1}\right)\right\}\exp(-i\omega t),\\ H_y&=2A_xn_1\left\{\cos\left(\frac{\omega z}{v_1}\right)\right\}\exp(-i\omega t),\quad H_z=0.\end{aligned}\right\}\tag{10}$$

Мы видим, что в каждый момент времени в первой среде фаза всюду постоянна и не существует конечной скорости распространения волн. Такие волны называются *стоячими*. Амплитуды электрического и магнитного векторов представляют собой периодические функции z. Плоскости нулевой амплитуды называются *узлами*, а плоскости, где амплитуды экстремальны,— *пучностями*. Из (9) находим, что узлы электрического поля будут при

$$z=\frac{m\pi v_1}{\omega}=\frac{m\lambda_0}{2n_1},\quad m=0,\ 1,\ 2,\ \ldots,\tag{11а}$$

а пучности — при

$$z=\frac{m\lambda_0}{2n_1},\quad m=\frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots\tag{11б}$$

Согласно (10) узлы магнитного поля совпадают с пучностями электрического поля, и наоборот. В частности, на отражающей поверхности находится узел электрического и пучность магнитного полей.

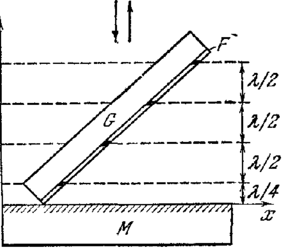


Рис. 7.18. Опыт Винера со стоячими волнами.
Угол между пластинкой и зеркалом значительно преувеличен

Существование стоячих световых волн было впервые установлено экспериментально Винером [12]. Его установка показана на рис. 7.18. Плоское зеркало M с наружным отражающим серебряным слоем освещалось нормально падающим параллельным пучком квазимонохроматического света. Тонкий слой F (толщиной меньше 1/20 длины волны света) прозрачной фотографической эмульсии, нанесенный на плоскую поверхность стеклянной пластинки G, помещался перед зеркалом M под малым углом к его поверхности. После проявления эмульсии были обнаружены черные эквидистантные полосы и прозрачные области между ними. Максимумы почернения располагались в местах пересечения эмульсии F с плоскостями пучностей как электрического, так и магнитного полей. Дальнейшие опыты с фотографической эмульсией, плотно прижатой к вогнутой зеркальной поверхности, показали, что почернение отсутствует в местах, непосредственно соприкасающихся с зеркалом. Отсюда Винер сделал вывод, что почерневшие места соответствуют пучностям электрического поля, так как на поверхности зеркала находится пучность магнитного поля *).

*) Здесь допускается, что уравнение (6), полученное для диэлектрика, справедливо и для поверхности, покрытой серебром. Теория отражения от металлов, разбираемая в гл. 13, показывает, что это справедливо и в условиях опытов Винера.

Следовательно, *фотохимическое действие прямо связано с электрическим, а не с магнитным вектором.* Такое заключение, конечно, вытекает из электронной теории. Фотографический процесс — это процесс ионизационный, при котором электрон удаляется из атомной связи галоидного серебра, а электромагнитная сила, действующая на заряженную частицу в покое, пропорциональна электрическому вектору (см. 1.1.34)).

Аналогичные эксперименты проводились с флуоресцирующими [13] и фотоэмиссионными пленками [14], служившими детекторами стоячих волн вместо фотографической эмульсии, применявшейся Винером. В обоих случаях, как и следовало ожидать на основании электронной теории, максимальный эффект обнаруживался в пучности электрического поля.

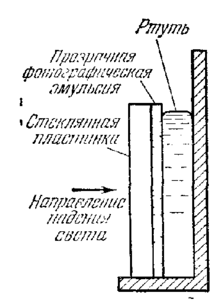


Рис 7.19. Схема устройства Липпмана для цветной фотографии.

Применение стоячих световых волн лежит в основе способа цветной фотографии, разработанного Липпманом [15]. Пластинка, покрытая прозрачной мелкозернистой фотоэмульсией, экспонируется в камере так, что эмульсия обращена в сторону, противоположную падающему свету. Непосредственно к эмульсии прилегает отражающий слой ртути (рис. 7.19). Предположим для простоты, что пластинка освещается падающим нормально квазимонохроматическим светом с длиной волны λ_0. Так как фотохимическое действие максимально в пучностях электрического поля (см. уравнение (11б)), то в проявленной пластинке серебро образует систему эквидистантных слоев, параллельных поверхности эмульсии, с оптическим расстоянием между ними, равным $\lambda_0/2$. Если теперь осветить пластинку падающим белым светом нормально к ее поверхности, то образовавшиеся слои серебра будут действовать как частично отражающие поверхности и отраженный свет будет состоять из ряда световых пучков с оптической разностью хода, равной целому числу, умноженному на λ_0. Ниже (см. § 7.6) мы рассмотрим интерференционную картину, возникающую в результате суперпозиции такого ряда световых пучков. Анализ показывает, что в этом случае при большом их числе наблюдается очень острый максимум результирующей интенсивности для длины волны λ_0. Липпмановская пластинка действует, следовательно, как селективный отражатель света с такой же длиной волны, какой она была освещена при изготовлении.

Винер [12] использовал свою установку и для исследования интерференционных явлений в линейно поляризованном свете при угле падения 45°. Он нашел, что при направлении электрических колебаний в падающем свете, перпендикулярном к плоскости падения, темные участки в эмульсии образуют систему эквидистантных параллельных полос: если же вектор электрических колебаний в падающем свете лежит в плоскости падения, то почернение оказывается равномерным. Этот результат снова подтверждает, что фотохимическое действие прямо связано с электрическим, а не магнитным полем. Когда направление электрических колебаний перпендикулярно к плоскости падения, то $A_\parallel=0$ и при $\theta_i=45°$ получаем из (7)

$$\left.\begin{aligned} &E_x=0,\\ &E_y=-2A_\perp\left\{\sin\left(\frac{\omega z}{\sqrt{2}v_1}\right)\right\}\exp\left\{-i\left[\omega\left(t-\frac{x}{\sqrt{2}v_1}\right)-\frac{\pi}{2}\right]\right\},\\ &E_z=0;\end{aligned}\right\} \qquad (12)$$

следовательно, амплитуда электрического вектора, а также усредненная по времени плотность электрической энергии изменяются периодически в направлении оси z. Если направление колебания лежит в плоскости падения, то $A_\perp=0$

и из (7) получим

$$\left.\begin{aligned} E_x &= \sqrt{2}\, A_\parallel \sin\left(\frac{\omega z}{\sqrt{2}\, v_1}\right) \exp\left\{-i\left[\omega\left(t - \frac{x}{\sqrt{2}\, v_1}\right) - \frac{\pi}{2}\right]\right\}, \\ E_y &= 0, \\ E_z &= -\sqrt{2}\, A_\parallel \cos\left(\frac{\omega z}{\sqrt{2}\, v_1}\right) \exp\left\{-i\left[\omega\left(t - \frac{x}{\sqrt{2}\, v_1}\right)\right]\right\}. \end{aligned}\right\} \quad (13)$$

Из (13) и (1.4.54) следует, что усредненная по времени плотность электрической энергии в этом случае равна

$$\langle w_e \rangle = \frac{n_1^2}{16\pi} \mathbf{E} \cdot \mathbf{E}^* = \frac{n_1^2}{16\pi} (E_x E_x^* + E_z E_z^*) = \frac{n_1^2}{8\pi} A_\parallel A_\parallel^* \quad (14)$$

и не зависит от z. Вместе с тем соотношение (8) указывает, что для магнитного поля положение оказывается обратным. Усредненная по времени плотность магнитной энергии периодически изменяется в направлении оси z, когда направление магнитных колебаний в падающем свете лежит в плоскости падения, и не зависит от z, если направление магнитных колебаний в падающем свете перпендикулярно к плоскости падения.

§ 7.5. Двухлучевая интерференция. Деление амплитуды

7.5.1. Полосы, получающиеся с плоскопараллельной пластинкой. Предположим, что плоскопараллельная пластинка из прозрачного материала освещается точечным источником S квазимонохроматического света (рис. 7.20). В любую точку P, находящуюся с той же стороны пластинки, что и S, приходят два луча — один, отразившийся от верхней поверхности пластинки, и другой, отразившийся от нижней ее поверхности. Интерферируя, эти лучи со стороны S образуют нелокализованную интерференционную картину. Из соображений симметрии очевидно, что полосы в плоскостях, параллельных пластинке, имеют вид колец с осью SN, нормальной к пластинке, и при любом положении P они перпендикулярны плоскости SNP. Из п. 7.3.4 следует, что видность этих интерференционных полос уменьшается при увеличении размеров источника в направлении, параллельном плоскости SNP. Случай, когда P находится в бесконечности и наблюдение ведется либо глазом, адаптированным на бесконечность, либо в фокальной плоскости объектива телескопа, служит важным исключением. В этих условиях оба луча, идущих от S к P, а именно лучи $SADP$ и $SABCEP$ (рис. 7.21), происходят от одного падающего луча и после прохождения пластинки параллельны. Оптическая разность хода между ними равна

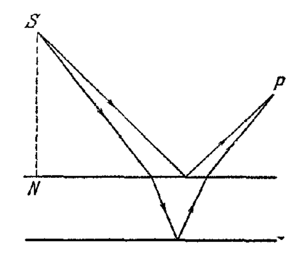


Рис. 7.20. Плоскопараллельная пластинка, освещенная точечным источником света.

$$\Delta\mathcal{S} = n'(AB + BC) - nAN, \quad (1)$$

где n' и n — показатели преломления пластинки и окружающей среды, а N — основание перпендикуляра, опущенного из C на AD. Если h — толщина пластинки, а θ и θ' — углы падения и преломления на верхней поверхности, то

$$AB = BC = \frac{h}{\cos\theta'}, \quad (2)$$

$$AN = AC \sin\theta = 2h \operatorname{tg}\theta' \sin\theta, \quad (3)$$

$$n' \sin\theta' = n \sin\theta. \quad (4)$$

Из (1), (2), (3) и (4) получаем

$$\Delta\mathcal{S} = 2n'h\cos\theta', \tag{5}$$

а соответствующая разность фаз равна

$$\delta = \frac{4\pi}{\lambda_0} n'h\cos\theta'. \tag{6}$$

Следует также учитывать изменение фазы на π, которое, согласно формулам Френеля (1.5.21а), происходит при каждом отражении от верхней или нижней поверхности. Полная разность фаз в P равна поэтому

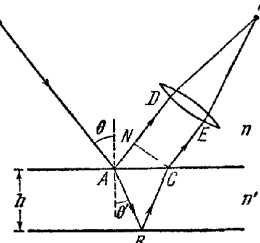


Рис. 7.21. Плоскопараллельная пластинка, возникновение интерференционных полос, локализованных в бесконечности.

$$\delta = \frac{4\pi}{\lambda_0} n'h\cos\theta' \pm \pi = \tag{7а}$$

$$= \frac{4\pi h}{\lambda_0}\sqrt{n'^2 - n^2\sin^2\theta} \pm \pi. \tag{7б}$$

Итак, θ определяется только положением точки P в фокальной плоскости телескопа, и следовательно, δ не зависит от положения источника S. Отсюда вытекает, что при использовании протяженного источника полосы оказываются столь же отчетливыми, как и с точечным источником. Так как это справедливо только для одной определенной плоскости наблюдения, то про такие полосы говорят, что они *локализованы*, а в данном случае — локализованы в бесконечности.

Интенсивность в интерференционной картине меняется в соответствии с соотношением (7.2.15), и из (7) и (7.2.16) находим, что светлые полосы расположены при

$$2n'h\cos\theta' \pm \frac{\lambda_0}{2} = m\lambda, \qquad m = 0,\ 1,\ 2,\ \ldots, \tag{8а}$$

а темные полосы — при

$$2n'h\cos\theta' \pm \frac{\lambda_0}{2} = m\lambda_0, \qquad m = \frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots \tag{8б}$$

Таким образом, заданная полоса характеризуется постоянством величины θ' (а значит, и θ) и, следовательно, создается светом, падающим на пластинку под каким-то определенным углом. Поэтому такие полосы часто называют *полосами равного наклона*. Если ось объектива телескопа нормальна к пластинке, то при нормальном отражении света ($\theta = \theta' = 0$) полосы имеют вид концентрических колец с центром в фокусе. Порядок интерференции максимален в центре картины, где его величина m_0 определяется соотношением

$$2n'h \pm \frac{\lambda_0}{2} = m_0\lambda_0; \tag{9}$$

m_0 не обязательно должно быть целым числом, и мы можем написать

$$m_0 = m_1 + e, \tag{10}$$

где m_1 — целое, кратное порядку внутренней наиболее светлой полосы, а e — меньше единицы и называется *дробным порядком в центре*. Для p-й от центра светлой полосы с угловым радиусом θ_p порядок интерференции m_p равен, в соответствии с (8а),

$$2n'h\cos\theta'_p \pm \frac{\lambda_0}{2} = m_p\lambda_0 = [m_1 - (p-1)]\lambda_0, \tag{11}$$

а из (9) — (11) находим

$$2n'h(1 - \cos\theta'_p) = (p - 1 + e)\lambda_0. \tag{12}$$

Итак, если θ_p' мало, то по закону преломления $n' \approx n\theta_p/\theta_p'$ и $1 - \cos\theta_p' \approx \theta_p'^2/2 \approx n^2\theta_p^2/2n'^2$; значит, согласно (12), имеем

$$\theta_p \approx \frac{1}{n}\sqrt{\frac{n'\lambda_0}{h}}\sqrt{p-1+e}. \qquad (13)$$

Таким образом, угловой масштаб наблюдаемой картины пропорционален $\sqrt{1/h}$, и если $e=0$, а значит, в ее центре имеется максимум интенсивности, то радиусы светлых полос пропорциональны квадратному корню из положительных целых чисел.

Аналогичные полосы, локализованные в бесконечности, можно получить с воздушной плоскопараллельной пластинкой, образованной внутренними плоскостями двух прозрачных пластин (рис. 7.22). При использовании такого устройства полосы наблюдают при постепенном изменении толщины h воздушной прослойки между пластинами. При увеличении h полосы возникают в центре картины, и при каждом увеличении h на $\lambda_0/2$ появляется новая полоса (предполагается, что для воздуха $n'=1$). Пластинки делают слегка клиновидными, для того чтобы избавиться от искажающих эффектов, связанных с отражением света на их внешних поверхностях. Оптическая разность хода лучей, идущих от S до P, для света, отраженного от внешних поверхностей, изменяется с положением S так, что в случае протяженного источника получается в среднем равномерное освещение в фокальной плоскости. Полосы, образованные светом, отраженным от внутренних параллельных поверхностей, налагаются на этот фон. Если пластинка не очень тонка, то полосы соответствуют высоким порядкам интерференции и поэтому не видны в белом свете. Например, при $h=1$ *см* и $n'=1{,}5$ порядок интерференции в центре составляет примерно 75 000 для $\lambda_0 = 4000$ Å и приблизительно 43 000 для $\lambda_0 = 7000$ Å; следовательно, в видимом спектре приблизительно 32 000 волн имеют максимум интенсивности в центре. Как мы видим из (5) и (7.3.15), для того чтобы полосы были отчетливы, требования к монохроматичности источника должны становиться тем строже, чем больше оптическая толщина пластинки. Ниже (см. п. 7.5.8) мы покажем, что существует верхний предел оптической толщины пластинки, при котором с данным источником полосы еще могут наблюдаться.

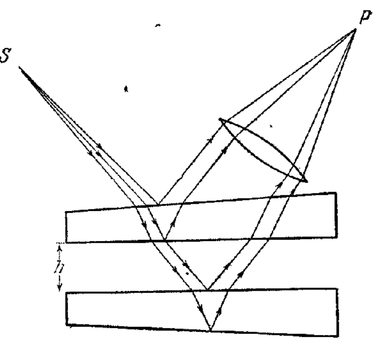


Рис. 7.22. Плоскопараллельная воздушная прослойка, возникновение полос, локализованных в бесконечности.

До сих пор мы предполагали, что оптическая толщина пластинки повсюду одинакова. Практически это предположение можно считать справедливым, если воспользоваться диафрагмой, ограничивающей освещенную площадь пластинки. Из соотношения (8) следует, что изменение оптической толщины пластинки на $\Delta(n'h)$ вызывает смещение картины на Δm порядков, где

$$\Delta m = \frac{2\cos\theta'}{\lambda}\Delta(n'h), \qquad (14)$$

а в центре, где $\theta'=0$

$$\Delta m = \frac{2}{\lambda_0}\Delta(n'h). \qquad (15)$$

Таким образом, при движении пластинки относительно диафрагмы изменения $n'h$ можно обнаружить по изменению порядка интерференции в центре. Подоб-

ный метод применяется в оптической промышленности для проверки качества пластин, от которых требуется постоянство оптической толщины [16].

Мы рассматривали пока только свет, отраженный от пластинки, но, конечно, подобные рассуждения применимы и для света, прошедшего сквозь пластинку. В этом случае (рис. 7.23) в точку P фокальной плоскости зрительной трубы приходят от источника S два луча: один, прошедший прямо, и другой — после двух внутренних отражений. Оптическую разность хода этих лучей находят таким же образом, как и при выводе (5), т. е.

$$\Delta\mathcal{S} = 2n'h \cos \theta', \tag{16}$$

а значит, соответствующая разность фаз равна

$$\delta = \frac{4\pi}{\lambda_0} n'h \cos \theta'. \tag{17}$$

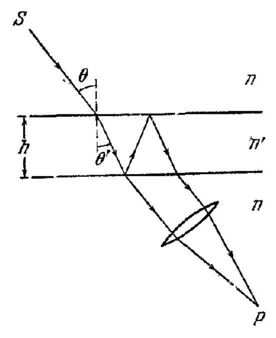


Рис. 7.23. Плоскопараллельная пластинка; возникновение полос в прошедшем свете, локализованных в бесконечности.

Дополнительная разность фаз, вызванная отражением, здесь отсутствует, так как оба внутренних отражения происходят в одинаковых условиях. Интерференционная картина, создаваемая протяженным источником, и в этом случае локализована в бесконечности. Сравнивая (17) и (7а), мы видим, что картины в проходящем и отраженном свете дополнительны, т. е. светлые полосы одной и темные полосы другой находятся на одном и том же угловом расстоянии относительно нормали к пластинке. Однако если отражательная способность поверхности пластинки мала (как, например, на границе стекло — воздух, где при нормальном падении она примерно равна 0,04), то интенсивности двух интерферирующих лучей, прошедших сквозь пластинку, очень сильно отличаются друг от друга. Поэтому (см. (7.2.16)) различие в интенсивности максимумов и минимумов оказывается малым, а видность полос — низкой.

Наше предыдущее рассуждение было не вполне строгим, так как мы пренебрегли многократностью внутренних отражений в пластинке. В действительности точку P достигает не два, как мы предположили, а целый ряд пучков, идущих от S. Но если отражение на поверхностях пластинки мало, то наше приближение вполне удовлетворительно, так как пучки после первых двух отражений обладают ничтожной энергией. Из более точного рассмотрения, приводимого далее (см. § 7.6), можно заключить, что положение максимумов и минимумов точно определяется соотношением (8), но при значительной отражательной способности поверхностей пластинки многократные отражения сильно изменят распределение интенсивности в полосах.

7.5.2. Интерференция в тонких пленках; интерферометр Физо. Допустим, что прозрачная пленка с плоскими, но не обязательно параллельными отражающими поверхностями освещается точечным источником S квазимонохроматического света. Два луча *), исходящие от S, а именно SAP и $SBCDP$ (рис. 7.24), приходят в какую-нибудь точку P с той же стороны пленки, где находится и S; следовательно, в этой области интерференционная картина не локализована. Оптическая разность хода между двумя путями от S до P равна

$$\Delta\mathcal{S} = n(SB + DP - SA - AP) + n'(BC + CD), \tag{18}$$

где n' и n — соответственно показатели преломления пленки и окружающей среды. Точную величину $\Delta\mathcal{S}$ трудно вычислить, но если пленка достаточно тон-

*) Мы снова пренебрегаем многократными отражениями; их влияние рассматривается ниже (см. § 7.6).

ка, то B, A и D находятся на очень малом расстоянии друг от друга, и значит,

$$nSA \approx nSB + n'BN_1, \qquad (19\text{ а})$$

а

$$nAP \approx nDP + n'N_2D, \qquad (19\text{ б})$$

где AN_1, AN_2 — перпендикуляры к BC и CD. Из (18) и (19) имеем

$$\Delta\mathscr{S} \approx n'(N_1C + CN_2). \qquad (20)$$

Кроме того, если угол между поверхностями пленки достаточно мал, то

$$N_1C + CN_2 \approx N'_1C + CN'_2. \qquad (21)$$

Здесь N'_1 и N'_2 — основания перпендикуляров, опущенных из E на BC и CD,

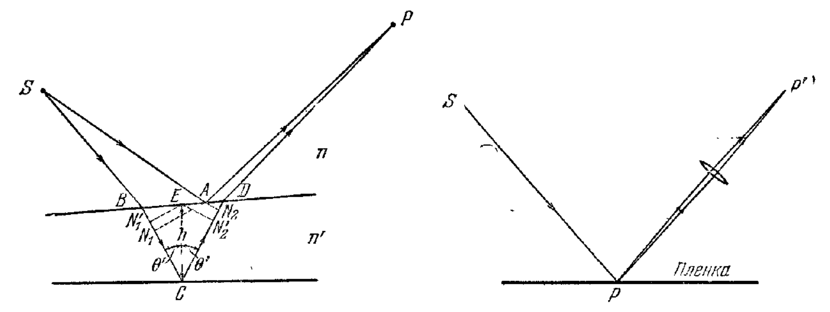


Рис. 7.24. Тонкая пленка, освещенная точечным источником света.

Рис. 7.25. Тонкая пленка; возникновение полос, локализованных в пленке.

а E — пересечение верхней поверхности с нормалью к нижней поверхности в точке C. Но

$$N'_1C = CN'_2 = h\cos\theta', \qquad (22)$$

где $h = CE$ — толщина пленки вблизи точки C, измеренная по нормали к нижней поверхности, θ' — угол отражения на внутренней поверхности пленки. Следовательно, для тонкой пленки, мало отличающейся от плоскопараллельной, можно написать, пользуясь (20), (21) и (22),

$$\Delta\mathscr{S} = 2n'h\cos\theta', \qquad (23)$$

а соответствующая разность фаз в P равна

$$\delta = \frac{4\pi}{\lambda_0} n'h\cos\theta'. \qquad (24)$$

В общем случае для данной точки P обе величины h и θ зависят от положения S и даже при небольшом увеличении размеров источника область значений δ в P становится столь большой, что полосы исчезают. Тем не менее существует специальный случай, когда P находится в пленке, а наблюдение ведется с микроскопом, сфокусированным на пленку, или сам глаз аккомодируется на нее. Тогда h практически одинаково для всех пар лучей от протяженного источника, приходящих в точку P', сопряженную с P (рис. 7.25), и различие величин δ в точке P' вызывается главным образом различием значений $\cos\theta'$. Если интервал изменений $\cos\theta'$ достаточно мал, то интервал величин δ в точке P' много меньше 2π, даже с источником значительных размеров, и полосы отчетливо видны. Очевидно, что они локализованы в пленке. Практически условие малости интервала изменений $\cos\theta'$ можно выполнить при наблюдении в направлении, близком к нормальному, или при ограничении входного зрачка, хотя зрачок невооруженного глаза и сам по себе может быть достаточно мал.

Учитывая изменение фазы на π при отражении на одной из поверхностей пленки, получим из (24) и (7.2.16), что максимумы интенсивности в P' (и, очевидно, в P) находятся при

$$2n'h\overline{\cos\theta'} \pm \frac{\lambda_0}{2} = m\lambda_0, \qquad m = 0,\ 1,\ 2,\ \ldots, \tag{25а}$$

и минимумы интенсивности — при

$$2n'h\overline{\cos\theta'} \pm \frac{\lambda_0}{2} = m\lambda_0, \qquad m = {}^1/_2,\ {}^3/_2,\ {}^5/_2,\ \ldots, \tag{25б}$$

где $\overline{\cos\theta'}$ — среднее значение $\cos\theta'$ для точек источника, от которых свет доходит в P'. Величина $n'h$, присутствующая в этих соотношениях, представляет собой оптическую толщину пленки в P, и если наше приближение остается в силе, интерференция в P не зависит от толщины пленки в других местах. Отсюда следует, что соотношения (25) остаются справедливыми даже при неплоских поверхностях пленки при условии, что угол между ними остается малым. Тогда, если $\overline{\cos\theta'}$ достаточно постоянен, интерференционные полосы соответствуют совокупности мест пленки, где оптические толщины одинаковы. По этой причине такие полосы часто называют *полосами равной толщины*

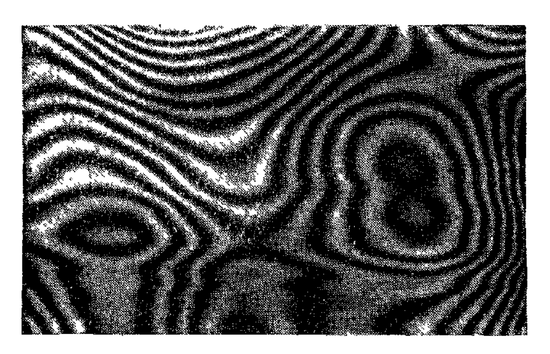


Рис. 7.26. Полосы равной толщины, полученные с тонкой стеклянной пластинкой.

(рис. 7.26). Их можно наблюдать в тонкой воздушной прослойке между отражающими поверхностями двух прозрачных пластинок. Вблизи нормального падения условия (25) для темной полосы при $\overline{\cos\theta'} = 1$ и длине волны в воздухе $\lambda = \lambda_0/n$ имеют вид

$$h = \frac{m\lambda}{2}, \qquad m = 0,\ 1,\ 2,\ \ldots \tag{26}$$

Таким образом, полосы вырисовывают контуры слоев равной толщины, причем расстояние между полосами соответствует изменению толщины на $\lambda/2$. Если толщина слоя повсюду постоянна, интенсивность по всей его поверхности одинакова. Это широко используется для испытания качества оптических поверхностей. При этом наблюдается прослойка воздуха между исследуемой поверхностью и поверхностью эталона (пробной пластинки) равной, но противоположной кривизны. При клиновидной воздушной прослойке между плоскими поверхностями полосы будут эквидистантны и параллельны ребру клина. Линейное расстояние между соседними светлыми полосами равно $\lambda/2\alpha$, где α — угол при вершине клина. Так, например, при $\alpha = 1'$ и $\lambda = 5500$ Å расстояние между полосами составляет примерно 1,9 *мм*; отсюда следует, что для достаточного разделения полос угол клина должен быть очень мал. Полосы, получающиеся с клином, используются для проверки концевых мер (калибров), служа-

щих стандартами длин в механических цехах. Испытуемый калибр G_1 (рис. 7.27) представляет собой стальную колодку с двумя параллельными, полированными плоскостями с определенным расстоянием между ними. Одну из этих поверхностей и одну из поверхностей эталона G_2 такой же номинальной длины плотно прижимают к плоской стальной поверхности; верхние поверхности обеих мер покрывают прозрачной пластинкой T с плоской поверхностью Обычно между пластинкой и мерами образуются клиновидные воздушные прослойки, в которых при освещении монохроматическим светом можно наблюдать интерференционные полосы Зная расстояние между полосами и между калибрами, можно определить разницу в длинах последних *).

Полосы, называемые *кольцами Ньютона* (рис. 7.28), служат еще одним примером полос равной толщины. Кольца Ньютона представляют исторический интерес в связи со взглядами самого Ньютона на природу света. Они наблюдаются в воздушном зазоре между соприкасающимися выпуклой сферической поверхностью линзы и плоской поверхностью стекла (рис. 7 29) Эти полосы имеют вид окружностей с центром в точке касания C Пусть R — радиус кривизны OC выпуклой поверхности, тогда, если пренебречь членами четвертого порядка, толщина зазора h на расстоянии r от OC равна

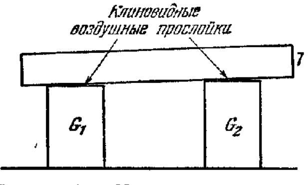


Рис 7 27 Интерферометрическое сравнение концевых мер

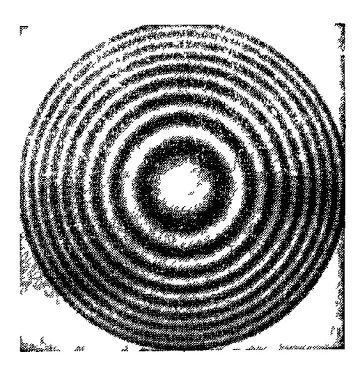


Рис. 7 28. Кольца Ньютона.

$$h = R - \sqrt{R^2 - r^2} \approx \frac{r^2}{2R}. \qquad (27)$$

При нормальном падении условия для радиусов темных полос, согласно (26) и (27), имеют вид

$$r = \sqrt{mR\lambda}, \qquad m = 0,\ 1,\ 2,\ \ldots, \qquad (28)$$

т. е радиусы темных полос пропорциональны квадратному корню из положительных целых чисел. При увеличении расстояния между линзой и пластинкой заданная толщина зазора h смещается по направлению к центру, причем кольца сжимаются и каждый раз при увеличении расстояния на $\lambda/2$ одно из них пропадает. Интересно отметить, что это устройство, так же как и опыт Юнга, позволяет очень простыми средствами приближенно определить длину волны света

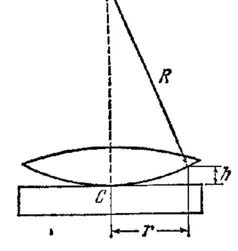


Рис 7 29. К возникновению колец Ньютона.

Если толщина пленки достигает всего лишь нескольких полуволн, порядки интерференции в монохроматической картине очень низки, и полосы становятся видимыми в белом свете. Примерами таких полос могут служить цвета мыльных

*) Изменение фазы при отражении от металлической поверхности не точно равно π, как считается в (25), но это не влияет на расстояние между полосами.

пузырей и масляных пятен на воде, видимые в отраженном свете. В устройстве для наблюдения колец Ньютона (см. рис. 7.29) в случае соприкосновения линзы и пластинки разность фаз в центре равна π для всех длин волн, и, следовательно, в этом месте в белом свете всегда будет темное пятно. По мере удаления от центрального пятна интерференционные кольца от различных монохроматических компонент света значительно расходятся, и при визуальном наблюдении вблизи центра видны цветные кольца. Последовательность цветов этих колец вполне определенна и известна как *цвета Ньютона*. Дальше от центра кольца различных порядков начинают перекрываться и освещение кажется глазу равномерно белым (см. § 7.3.3). Аналогично в клиновидной воздушной прослойке в белом свете у ребра клина отчетливо может наблюдаться темная полоса.

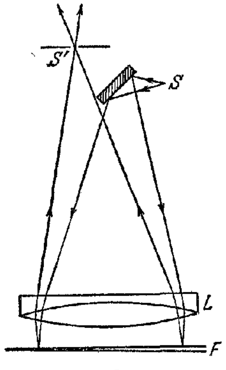


Рис. 7.30. Схема интерферометра Физо.

До сих пор мы говорили только об отраженном свете, однако интерференционная картина, локализованная в пленке, видна также и в проходящем свете. Как и в случае плоскопараллельной пластинки, картины в отраженном и прошедшем свете дополнительны, т. е. светлые полосы одной появляются в тех же местах пленки, что и темные полосы другой. При использовании мало отражающих поверхностей полосы в проходящем свете плохо видны вследствие неравенства интенсивностей интерферирующих пучков.

Мы показали, что необходимым условием отчетливости полос служит ограниченность диапазона величин $\cos\theta'$, соответствующих каждой точке пленки, и что полосы вырисовывают контуры слоев равной оптической толщины, только если $\overline{\cos\theta'}\approx 1$. Оба условия одновременно выполняются на большой площади пленки в *интерферометре Физо* [17] (рис. 7.30). Пучок света от квазимонохроматического источника S после отражения от маленького зеркала, превращается линзой L в параллельный и падает почти нормально на пленку F. Свет, отраженный поверхностями пленки, снова проходит через линзу L и собирается в отверстие S' в фокальной плоскости линзы L. Глаз, помещенный сразу за S' и сфокусированный на пленку, видит полосы, вырисовывающие контуры слоев равной оптической толщины на всей поверхности пленки, освещенной линзой L. Эти полосы часто называют *полосами Физо*. Как мы увидим (см. п. 7.5.3), полосы Физо можно получить и в толстых пленках при условии, что источник достаточно мал. Такие интерферометры применяются в оптических мастерских для проверки постоянства оптических толщин плоскопараллельных прозрачных пластинок. В Англии они используются в Национальной физической лаборатории для измерения концевых мер или калибров [18]. Устройство такого прибора показано на рис. 7.31, *а*. Одна из поверхностей калибра G плотно прижата к плоской полированной поверхности стальной плиты B. Верхнюю поверхность G, параллельную плите B, покрывают прозрачной пластинкой T с плоской нижней поверхностью. Вообще говоря, между T и G и между T и B образуются клиновидные воздушные слои и в каждом из них может наблюдаться интерференционная картина, состоящая из прямых параллельных линий с одинаковым расстоянием между ними в обеих картинах. Изменяя наклон T, можно установить полосы перпендикулярно к одному краю калибра (рис. 7.31, *б*). Пусть в какой-то точке на краю калибра толщины воздушных клиньев над G и B составляют соответственно h_1 и h_2, а соответствующие порядки интерференции равны m_1 и m_2; тогда, согласно (26), длина d калибра равна

$$d=h_2-h_1=(m_2-m_1)\frac{\lambda}{2}=\Delta m\frac{\lambda}{2}, \qquad (29)$$

где Δm — смещение порядка обеих картин. Таким образом, мы можем написать

$$d=(x+e)\frac{\lambda}{2}, \tag{30}$$

где x — неизвестное целое число, а $e<1$ — дробная часть порядка, которую можно определить интерполяцией. Величину x находят *методом дробных*

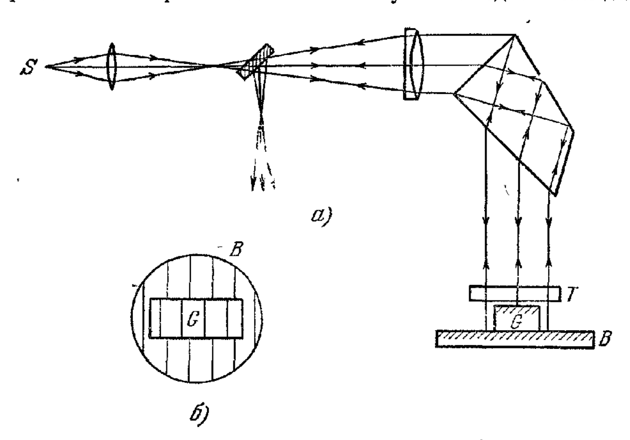


Рис. 7.31. Интерферометр для измерения концевых мер, применяемый в английской Национальной физической лаборатории.
а — схема установки, *б* — поле зрения.

частей порядка, который был впервые применен в интерферометрии Бенуа [19]. Дробные части порядка измеряются в свете четырех известных волн, которые (при использовании подходящего источника) удобно выделять по очереди вращением призмы постоянного отклонения D. Тогда мы получим соотношения

$$d=(x_1+e_1)\frac{\lambda_1}{2}=(x_2+e_2)\frac{\lambda_2}{2}=(x_3+e_3)\frac{\lambda_3}{2}=(x_4+e_4)\frac{\lambda_4}{2}, \tag{31}$$

где x_1, x_2, x_3 и x_4 — неизвестные целые числа; эти соотношения определяют «четверки» возможных значений величин x_1, x_2, x_3 и x_4. Измеряя приближенно микрометром длину d, можно решить, какая из таких «четверок» правильна, и, следовательно, получить точную величину d. С надлежащими предосторожностями можно измерять калибры длиной до 10 *см*. Если дробная часть порядка определена с точностью до 0,1, то для $\lambda=5000$ Å точность в измерении калибра составит $\pm 2{,}5\cdot 10^{-6}$ *см*.

Очевидно, что для устройства Физо существует предельный случай, когда источник S (см. рис. 7.30) уменьшается до точки и, значит, θ' принимает одинаковые значения во всех точках пленки. В этих условиях, однако, если поверхность пленки плоская, полосы должны быть нелокализованными, т. е. в любой плоскости, в которой встречаются отраженные пучки, они будут так же отчетливы, как и в пленке. Мы, следовательно, вынуждены более тщательно выяснить вопрос о локализации полос и о ее связи с размерами источника.

7.5.3. Локализация полос*). Интерференционные устройства, рассмотренные до сих пор, можно в общем охарактеризовать как устройства, в которых свет от источника достигает точек в области интерференции двумя различными путями. Пусть P будет такой точкой, и предположим, что свет длиной волны

*) Более полное рассмотрение вопроса о локализации полос приведено в [20].

λ_0 испускается квазимонохроматическим точечным источником S. Если SA_1B_1P и SA_2B_2P — два луча от S до P (рис. 7.32), то разность фаз в точке P равна

$$\delta_0 = \frac{2\pi}{\lambda_0}\{[SA_2B_2P] - [SA_1B_1P]\}. \qquad (32)$$

Величина δ_0 зависит от положения P, но она однозначно определена для всех P, так что интерференционные полосы, являющиеся геометрическим местом точек, для которых δ_0 постоянна, образуются в любой плоскости той области, где встречаются оба луча от S. Мы говорим, что такие полосы *не локализованы*. Они всегда наблюдаются с точечным источником, и их видность зависит только от относительной интенсивности обеих волн.

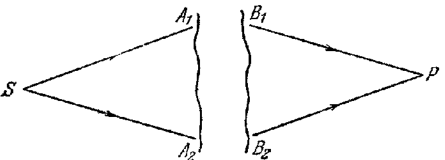


Рис 7.32. К интерференции двух пучков света от точечного источника.

Допустим теперь, что квазимонохроматический первичный источник занимает некоторое пространство вокруг S. Как и в п. 7.3.4, предположим, что такой источник состоит из некогерентных точечных источников, каждый из которых создает нелокализованную интерференционную картину. Тогда в каждой точке полная интенсивность равна сумме интенсивностей таких элементарных картин. Если в точке P разность фаз излучения от различных точек протяженного источника неодинакова, то элементарные картины смещены друг относительно друга в окрестности P и видность полос в точке P меньше, чем в случае точечного источника. Вообще говоря, как будет видно из дальнейшего, взаимное смещение (а следовательно, и уменьшение видности) растет по мере увеличения источника, но зависит от положения P. Таким образом, хотя мы имеем здесь дело с протяженным источником, видность полос в некоторых точках P может остаться такой же (или почти такой же), как и в случае точечного источника, тогда как в другом месте она упадет практически до нуля. Такие полосы характерны для протяженного источника и называются *локализованными*.

Уменьшение видности, связанное с заданными размерами источника, вообще говоря, трудно поддается расчету, так как оно зависит не только от относительной интенсивности элементарных картин, но и от их взаимного смещения. Однако здесь легко выделить два крайних случая. Взаимное смещение элементарных картин, малое по сравнению с одним порядком, очевидно, незначительно влияет на полосы, и при определенных обстоятельствах (см. рис. 7.11) уменьшение видности остается незаметным, по крайней мере при визуальном наблюдении, пока

$$\delta_{\text{макс}} - \delta_{\text{мин}} \leqslant \frac{\pi}{2}. \qquad (33)$$

Здесь $\delta_{\text{макс}}$ и $\delta_{\text{мин}}$ — максимальная и минимальная разность фаз излучения в точке P, приходящего от различных точек протяженного источника. Этот критерий справедлив также для источника с рассмотренным ранее распределением интенсивности (см. рис. 7.15, *б*, *в*), и мы вправе предположить, что он всегда выполняется. Вместе с тем, когда

$$\delta_{\text{макс}} - \delta_{\text{мин}} \gg \pi, \qquad (34)$$

взаимное смещение элементарных картин достигает многих порядков, и видность полос становится очень малой.

Исследуем теперь зависимость δ для некоторой точки S' протяженного источника от ее положения относительно точки S. Точки S и P являются «сопряженными» в том смысле, что от точечного источника с длиной волны λ_0,

находящегося в P, в точку S пришли бы две волны с разностью фаз δ_0. Пусть W_1 и W_2— волновые фронты этих двух волн, проходящие через точку S и нормальные к SA_1 и SA_2 соответственно (рис. 7.33). Их кривизны зависят от положений S и P и от оптических особенностей данного устройства. Пусть нормали, опущенные из точки S' на волновые фронты W_1 и W_2, пересекаются с ними в точках N_1 и N_2 соответственно. Оптическая длина пути от N_1 до P равна $[SA_1B_1P]$, а оптическая длина пути от N_2 до P равна $[SA_2B_2P]$. Разность фаз в точке P, соответствующей S', составляет, следовательно,

$$\delta = \frac{2\pi}{\lambda_0}\{[S'N_2] + [SA_2B_2P] - [S'N_1] - [SA_1B_1P]\}, \tag{35}$$

или, учитывая (32),

$$\delta - \delta_0 = \frac{2\pi n}{\lambda_0}(s_2 - s_1), \tag{36}$$

где $s_1 = S'N_1$, $s_2 = S'N_2$ и n — показатель преломления среды, окружающей источник.

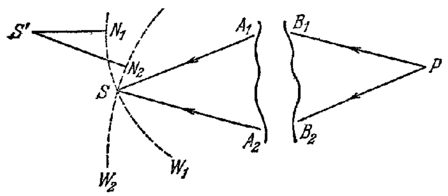


Рис. 7.33. К интерференции двух пучков света от протяженного источника.

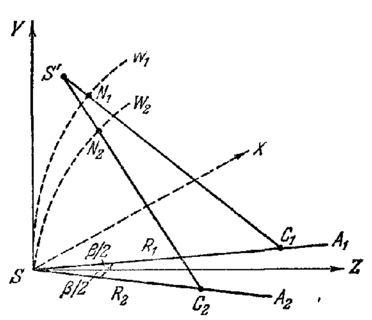


Рис. 7.34. К исследованию локализации полос в двухлучевом интерференционном устройстве.

Пусть S — начало прямоугольных координат с осями SX, SY и SZ, причем SZ и SX — соответственно внутренняя и внешняя биссектрисы угла A_1SA_2 (рис. 7.34). Пусть далее радиусы кривизны волновых фронтов W_1 и W_2 в плоскостях $S'SA_1$ и $S'SA_2$ равны соответственно R_1 и R_2; тогда центрами их кривизны C_1 и C_2 служат точки $\{R_1\sin(\beta/2),\ 0,\ R_1\cos(\beta/2)\}$ и $\{-R_2\sin(\beta/2), 0, R_2\cos(\beta/2)\}$ соответственно, где β — угол A_1SA_2. Если (x, y, z) — координаты точки S', то

$$C_1S' = s_1 + R_1 = \sqrt{\left(x - R_1\sin\frac{\beta}{2}\right)^2 + y^2 + \left(z - R_1\cos\frac{\beta}{2}\right)^2} = \\ = R_1\sqrt{1 - \frac{2}{R_1}\left(x\sin\frac{\beta}{2} + z\cos\frac{\beta}{2}\right) + \frac{(x^2+y^2+z^2)}{R_1^2}}. \tag{37}$$

Если размеры источника малы по сравнению с R_1, мы можем пренебречь степенями x/R_1, y/R_1 и z/R_1 выше второй; тогда

$$s_1 \approx -x\sin\frac{\beta}{2} - z\cos\frac{\beta}{2} + \frac{1}{2R_1}\left\{x^2 + y^2 + z^2 - \left(x\sin\frac{\beta}{2} + z\cos\frac{\beta}{2}\right)^2\right\}. \tag{38}$$

Аналогично

$$s_2 \approx x\sin\frac{\beta}{2} - z\cos\frac{\beta}{2} + \frac{1}{2R_2}\left\{x^2 + y^2 + z^2 - \left(x\sin\frac{\beta}{2} - z\cos\frac{\beta}{2}\right)^2\right\}. \tag{39}$$

В большинстве случаев, представляющих практический интерес, β мало, и можно отбросить члены в разложениях (38) и (39), содержащие β^2 и $\beta z/R$. Тогда из (36), (38) и (39) мы получим

$$\delta - \delta_0 \approx \frac{2\pi n}{\lambda_0}\left\{\beta x + \frac{1}{2}\left(\frac{1}{R_2} - \frac{1}{R_1}\right)(x^2 + y^2)\right\}. \tag{40}$$

Если здесь можно пренебречь членом, не зависящим от β, то

$$\delta - \delta_0 \approx \frac{2\pi n}{\lambda_0}\beta x. \tag{41}$$

В таком приближении $\delta = \delta_0$ при $x = 0$, так что разность фаз (и, следовательно, уменьшение видности) незначительна, если мы пользуемся источником, протяженным в плоскости YSZ. Согласно (33) и (41) уменьшение видности неощутимо, если размеры источника в направлении, нормальном к этой плоскости, не превышают $\lambda_0/4n\beta$. Итак, направления SA_1 и SA_2 и, следовательно, направления выбранных координатных осей, вообще говоря, зависят от положения P. Если для всех точек P плоскости YSZ имеют общую линию пересечения, то линейный источник (практически щель шириной, меньшей $\lambda_0/4n\beta$), вытянутый вдоль этой линии, дает в любой точке полосы такой же видности, как и точечный источник; полосы остаются нелокализованными. Все это относится ко всем случаям, описанным в п. 7.3.2 (исключая опыт Меслина). В случае зеркал Френеля, например, при любом положении точки P линия, проходящая через S и параллельная линии пересечения зеркал, лежит в плоскости YSZ, и при линейном источнике (щель), ориентированном также вдоль этой линии, наблюдаются, как было установлено в п. 7.3.4, нелокализованные полосы. Легко показать, что выражение (7.3.26) для ширины щели эквивалентно условию $e \ll \lambda_0/4n\beta$. Однако, вообще говоря, плоскости YSZ для всех P не имеют общего пересечения. Для получения картин в точках P линейный источник должен находиться только в ограниченной области плоскости YSZ. Если же протяженный источник занимает область, границы которой отстоят от этой плоскости на расстоянии, большом по сравнению с $\lambda_0/2n\beta$, то, согласно (34) и (41), видность будет очень мала. При этих обстоятельствах полосы от линейного источника будут локализованы в области, положение которой зависит от ориентации источника.

Величина β зависит от положения P, и если мы пренебрежем зависимостью R_1 и R_2 от β, то, согласно (40), величина $\delta - \delta_0$ для данного x уменьшается при уменьшении β. Отсюда следует, что если размеры источника увеличиваются в направлении, нормальном к плоскости YSZ, то полосы локализуются в области, соответствующей достаточно малым значениям β. В частности, $\beta = 0$, если SA_1 и SA_2 совпадают; в области локализации лежат и те точки P (если они вообще существуют), которые находятся на пересечении обоих лучей, образованных из одного падающего луча, выходящего из S. В окрестности этих точек β пренебрежимо мало, и мы получим из (40)

$$\delta - \delta_0 \approx \frac{\pi n}{\lambda_0}\left(\frac{1}{R_2} - \frac{1}{R_1}\right)(x^2 + y^2). \tag{42}$$

Пусть W_1 и W_2— сферические волновые фронты; тогда R_1 и R_2 не зависят от x и y, и уменьшение видности не зависит от направления, в котором источник вытянут в плоскости YSZ. Для источника в виде диска, лежащего в этой же плоскости, уменьшение видности, согласно (33) и (42), незначительно при условии, что его радиус не превышает

$$\sqrt{\lambda_0 R_1 R_2 / (2n\,|R_1 - R_2|)}.$$

В случае плоскопараллельной пластинки, рассматриваемой в зрительную трубу (см. п. 7.5.1), $\beta = 0$, когда точка P находится в фокальной плоскости объектива, а R_1 и R_2 бесконечны при всех положениях S, так что в соответствии с (42) $\delta = \delta_0$ для любых x и y. Следовательно, в фокальной плоскости объектива видность полос не зависит ни от положения, ни от размеров источника. На практике апертура объектива всегда ограничивает эффективный размер источника. Следовательно, при наблюдении полос вне фокальной плоскости объектива (как, например, при неправильной установке зрительной трубы) уменьшение видности остается незначительным до тех пор, пока эффективный размер источника не превысит величины, допускаемой (33) и (40).

В качестве примера рассмотрим полосы в отраженном свете от прозрачного клина с небольшим углом α; пусть его показатель преломления равен n', а сам он ограничен плоскими поверхностями и находится в среде с показателем

преломления *) n. Пусть S — точечный источник квазимонохроматического света. Рассмотрим луч, лежащий в главном сечении клина (сечение под прямым углом к ребру) и падающий на переднюю поверхность клина в точке A (рис. 7.35). Этот луч вызывает появление отраженного луча AP и преломленного AB. Преломленный луч AB после отражения в B от задней поверхности клина

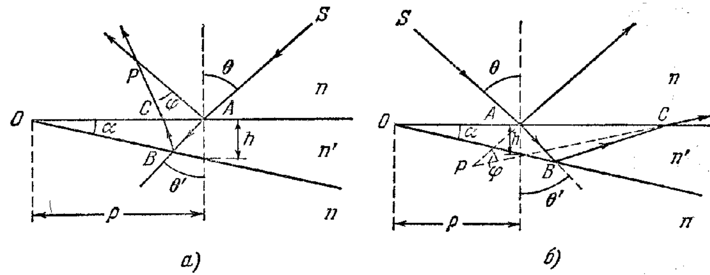


Рис. 7.35. Лучи, отраженные и преломленные на поверхностях плоского клина.

и преломления в C на передней его поверхности, выходит в направлении CP и встречает отраженный луч AP в точке P. Если источник занимает конечную область вокруг S, то полосы вблизи плоскости SAP локализованы в окрестности P.

Пусть ρ — расстояние между A и ребром клина O, θ и θ' — соответственно углы падения и преломления в точке A, а φ — угол APC. Если B и O находятся по одну сторону от нормали, восстановленной в A, к задней поверхности клина, то точка P оказывается действительной (рис. 7.35, a): угол отражения в B равен $(\theta'-\alpha)$, угол выхода луча в точке C равен $(\theta-\varphi)$, и из элементарной геометрии следует, что

$$AB=\frac{\rho\sin\alpha}{\cos(\theta'-\alpha)}, \tag{43}$$

$$\frac{BC}{AB}=\frac{\cos\theta'}{\cos(\theta'-2\alpha)}, \tag{44}$$

$$\frac{AC}{AB}=\frac{\sin 2(\theta'-\alpha)}{\cos(\theta'-2\alpha)}, \tag{45}$$

$$\frac{AP}{AC}=\frac{\cos(\theta-\varphi)}{\sin\varphi}, \tag{46}$$

$$\frac{CP}{AC}=\frac{\cos\theta}{\sin\varphi}. \tag{47}$$

По закону Снеллиуса для преломления в точках A и C получим

$$n'\sin\theta'=n\sin\theta, \tag{48}$$

$$n'\sin(\theta'-2\alpha)=n\sin(\theta-\varphi). \tag{49}$$

Вычитая одно из другого и пользуясь тождеством

$$\sin a-\sin b=2\cos\frac{a+b}{2}\sin\frac{a-b}{2},$$

находим

$$\sin\frac{\varphi}{2}=\frac{n'\sin\alpha\cos(\theta'-\alpha)}{n\cos\left(\theta-\frac{\varphi}{2}\right)}, \tag{50}$$

и, если величина θ не слишком близка к $\pi/2$, мы можем написать для малого α

$$\varphi\approx\frac{2n'\alpha\cos\theta'}{n\cos\theta}. \tag{51}$$

*) Более полно этот вопрос рассмотрен в [21].

Из (43), (44) и (46) имеем

$$AP = \frac{2\rho \sin\alpha \sin(\theta' - \alpha)\cos(\theta - \varphi)}{\cos(\theta' - 2\alpha)\sin\varphi}. \tag{52}$$

При малом α последнее соотношение можно разложить по степеням α, используя (51). Оставляя только главный член и принимая во внимание (48), получим

$$AP \approx \frac{\rho n^2 \sin\theta \cos^2\theta}{n'^2 - n^2 \sin^2\theta}. \tag{53}$$

Члены, отброшенные в (53), порядка $\rho\alpha$, т. е. порядка толщины клина в точке A. Если B и O находятся по разные стороны от нормали к задней поверхности клина, восстановленной в точке A (рис. 7.35, *б*), то точка P оказывается мнимой, но приближенная формула (53) остается справедливой.

Пусть клин представляет собой воздушную прослойку между двумя стеклянными пластинками и мы вправе пренебречь эффектами, вызванными преломлением в передней пластинке; тогда можно положить в (53) $n = n' = 1$, $AP \approx \rho \sin\theta$ и угол $OPA \approx \pi/2$. Геометрическое место точек P имеет тогда вид

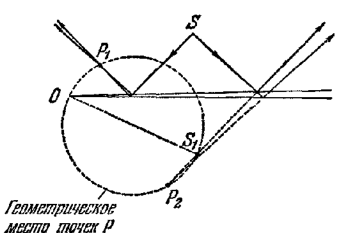


Рис. 7.36. К локализации полос в главном сечении воздушного клина.
Преломление на поверхностях клина не учитывается

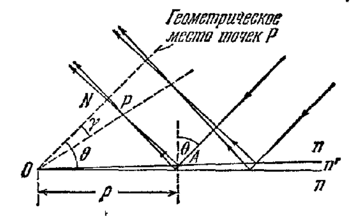


Рис. 7.37. К локализации полос, получающихся с клином, от источника света, находящегося в бесконечности.

окружности с диаметром OS_1, где S_1— зеркальное изображение S в передней поверхности клина (рис. 7.36).

Если источник S находится так далеко от клина, что все падающие лучи можно считать параллельными *) SA, то 0, а значит, θ' и φ не зависят от ρ, и из (52) следует, что AP пропорционально ρ. В этом случае геометрическое место точек P — плоскость, в которой лежит ребро клина (рис. 7.37). Угол γ между этой плоскостью локализации и плоскостью, нормальной к отраженному пучку, определяется из (53) и соотношения $AN = \rho \sin\theta$ в виде

$$\operatorname{tg}\gamma = \frac{AN - AP}{ON} \approx \left(\frac{n'^2 - n^2}{n'^2 - n^2 \sin^2\theta}\right) \operatorname{tg}\theta. \tag{54}$$

Очевидно, что $\gamma \approx 0$, если $\theta = 0$, и при нормальном падении плоскость локализации полос совпадает с передней поверхностью клина.

Рассмотрим далее оптическую разность хода $\Delta\mathscr{S}$ в точке P. Из рис. 7.35, *а* для действительной точки P имеем

$$\Delta\mathscr{S} = n'(AB + BC) + n(CP - AP). \tag{55}$$

Согласно (43) и (44) для малого α и величины θ', не слишком близкой к $\pi/2$, находим

$$AB + BC = \frac{2\rho \sin\alpha \cos\alpha}{\cos(\theta' - 2\alpha)} \approx \frac{2h}{\cos\theta'}, \tag{56}$$

где

$$h = \rho \operatorname{tg}\alpha \tag{57}$$

*) Это условие полностью удовлетворяется в устройстве Физо (см. п. 7.5.2).

— толщина клина в точке A. В этом приближении мы отбрасывали члены со степенями α. Аналогичным образом из (43), (45), (46) и (47) получим

$$CP-AP=\frac{-2\rho\sin\alpha\sin\left(\theta-\frac{\varphi}{2}\right)\sin(\theta'-\alpha)}{\cos\frac{\varphi}{2}\cos(\theta'-2\alpha)}\approx-\frac{2h\sin\theta\sin\theta'}{\cos\theta'}. \quad (58)$$

Следовательно, из (55), (56), (58) и (48) имеем

$$\Delta\mathscr{S}\approx 2n'h\cos\theta', \quad (59)$$

и соответствующая разность фаз в точке P равна

$$\delta\approx\frac{4\pi}{\lambda_0}n'h\cos\theta'=\frac{4\pi h}{\lambda_0}\sqrt{n'^2-n^2\sin^2\theta}. \quad (60)$$

Подобным же способом мы найдем, что это приближение справедливо и для условий, показанных на рис. 7.35, *б*, когда P — мнимая точка. Если учесть изменение фазы на π, происходящее при отражении на одной из поверхностей клина, то в соответствии с (60) и (7.2.16) находим, что максимумы интенсивности в P будут при

$$2n'h\cos\theta'\pm\frac{\lambda_0}{2}=m\lambda_0, \qquad m=0,\ 1,\ 2,\ \ldots, \quad (61а)$$

а минимумы интенсивности — при

$$2n'h\cos\theta'\pm\frac{\lambda_0}{2}=m\lambda_0,$$
$$m=\frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots \quad (61б)$$

Очевидно, что при постоянном угле θ' полосы в плоскости локализации эквидистантны и параллельны ребру клина.

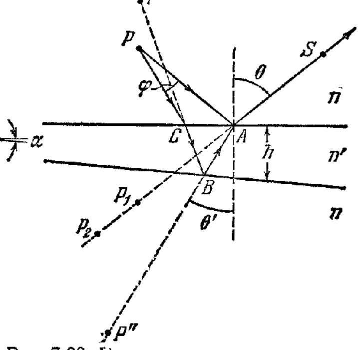


Рис. 7.38. Положение центров кривизны волновых фронтов от точечного источника света после отражения и преломления на поверхностях плоского клина.

Наконец, определим, как велики могут быть размеры источника, окружающего точку S, при которых еще нет заметного уменьшения видности полос в P. Представим себе, что точечный источник помещен в точку P, и определим радиусы кривизны R_1 и R_2 волновых фронтов W_1 и W_2, достигающих S после отражений на передней и задней поверхностях клина (рис. 7.38). Центр кривизны сферической поверхности W_1 лежит в P_1 на AS, где точка P_1 — зеркальное изображение P в передней поверхности клина. Таким образом,

$$R_1=SP_1=SA+AP. \quad (62)$$

Волновой фронт W_2, вообще говоря, отличен от сферического. Рассмотрим сначала его радиус кривизны в главном сечении клина. После преломления в C центр кривизны P' лежит на продолжении BC и его положение определяется формулой (4.6.21). В обозначениях, принятых в § 4.6, имеем $n_0=n$, $n_1=n'$, $r_y=\infty$, $\theta_0=\theta-\varphi$, $\theta_1=\theta'-2\alpha$, $d_0^{(t)}=CP$, $d_1^{(t)}=CP'$ и из соотношения (4.6.21) получаем

$$\frac{CP'}{CP}=\frac{n'\cos^2(\theta'-2\alpha)}{n\cos^2(\theta-\varphi)}. \quad (63)$$

После отражения в B центр кривизны окажется в точке P'' на продолжении AB, где P'' — зеркальное изображение P' в задней поверхности клина; после преломления в A центр кривизны окажется в точке P_2 на AS и, согласно (4.6.21)

(при $n_0=n'$, $n_1=n$, $r_y=\infty$, $\theta_0=\theta'$, $\theta_1=\theta$, $d_0^{(t)}=AP''$, $d_1^{(t)}=AP$), находим

$$\frac{AP_2}{AP''}=\frac{n\cos^2\theta}{n'\cos^2\theta'}. \tag{64}$$

Радиус кривизны R_2 фронта W_2 в главном сечении клина равен

$$R_2=SP_2=SA+AP_2, \tag{65}$$

где, согласно (64), и соотношению $AP''=AB+BC+CP'$

$$AP_2=\frac{n\cos^2\theta}{n'\cos^2\theta'}(AB+BC+CP'), \tag{66}$$

или, используя (63),

$$AP_2=\frac{n\cos^2\theta}{n'\cos^2\theta'}(AB+BC)+\frac{\cos^2\theta\cos^2(\theta'-2\alpha)}{\cos^2\theta'\cos^2(\theta-\varphi)}CP. \tag{67}$$

Для малого α имеем

$$AP_2\approx\frac{n\cos^2\theta}{n'\cos^2\theta'}(AB+BC)+CP\approx\frac{n\cos^2\theta}{n'\cos^2\theta'}\cdot\frac{2h}{\cos\theta'}+AP-\frac{2h\sin\theta\sin\theta'}{\cos\theta'}. \tag{68}$$

Здесь мы воспользовались соотношениями (56) и (58). Из (62), (65) и (68) получим окончательно

$$R_2-R_1\approx\frac{2h}{\cos\theta'}\left(\frac{n\cos^2\theta}{n'\cos^2\theta'}-\sin\theta\sin\theta'\right). \tag{69}$$

Аналогично можно вывести соответствующие выражения для сечения W_2, перпендикулярного главному сечению клина. Вместо (4.6.21) нужно пользоваться (4.6.22), и тогда вместо (63) и (64) мы получим

$$\frac{CP'}{CP}=\frac{n'}{n},\qquad \frac{AP_2}{AP''}=\frac{n}{n'}. \tag{70}$$

Это приводит к следующей формуле для разности радиусов кривизны сечений W_1 и W_2, перпендикулярных к главному сечению клина:

$$R_2-R_1\approx\frac{2h}{\cos\theta'}\left(\frac{n}{n'}-\sin\theta\sin\theta'\right). \tag{71}$$

Как уже отмечалось раньше, из (33) и (42) можно заключить, что уменьшение видности в точке P незначительно, если радиальные размеры источника относительно SA не превышают $\sqrt{\lambda_0R_1R_2/(2n\,|R_1-R_2|)}$. Из (69) и (71) следует, что при θ, не слишком близком к $\pi/2$, величина R_2-R_1 будет порядка h, и если R_1 достаточно велико по сравнению с h, как в случае устройства Физо, то можно принять $R_2/R_1\sim 1$. В этом приближении допустимый *угловой* радиус источника, наблюдаемый из точки P_1, равен

$$\varepsilon\approx\frac{1}{R_1}\sqrt{\frac{\lambda_0R_1R_2}{2n\,|R_1-R_2|}}\approx\sqrt{\frac{\lambda_0}{2n\,|R_1-R_2|}}. \tag{72}$$

В частности, для почти нормального падения света из (69) и (71) следует, что $R_2-R_1\approx 2hn/n'$, и (72) принимает вид

$$\varepsilon\approx\frac{1}{2n}\sqrt{\frac{\lambda_0n'}{h}}. \tag{73}$$

Для типичных значений для тонкого воздушного клина $h=0{,}01$ *см*, $n'=n\sim 1$, $\lambda_0=5500$ Å выражение (73) дает $\varepsilon\sim 2°$. Очевидно, допустимые размеры источника пропорциональны $\sqrt{1/h}$. Например, в только что разобранном случае при $h=1$ *см* величина ε, согласно (73), примерно равна 12′. Поэтому при исследовании интерферометра Физо мы говорили, что можно получить полосы с большой видностью с толстым клином, если источник достаточно мал.

7.5.4. Интерферометр Майкельсона. В устройствах, описанных в пп 7.5.1 и 7.5.2 суперпозиция обоих интерферирующих пучков происходит всюду.

кроме пространства, ограниченного поверхностями пластинки или пленки. Иногда это обстоятельство оказывается неудобным, и его можно устранить с помощью вспомогательной полуотражающей поверхности, на которой пучки полностью разделяются. Затем пучки опять сводятся. На этом принципе основано устройство *интерферометра Майкельсона* [22].

Простейшая схема такого прибора показана на рис. 7.39. Свет от протяженного источника S делится полупрозрачной поверхностью $\mathcal{A}$ плоскопараллельной стеклянной пластинки D на два пучка, направленных под прямым углом друг к другу. Отразившись от плоских зеркал M_1 и M_2, они возвращаются к D, где вновь соединяются и попадают в зрительную трубу T. Зеркало M_2 неподвижно, тогда как зеркало M_1, установленное на столике, можно передвигать микрометрическим винтом в направлении пластинки D или от нее. Пучок, отраженный от M_1, проходит сквозь диспергирующую пластинку D три раза, прежде чем попадает в T. Пучок, отраженный от M_2, проходит через пластинку D только один раз. Чтобы устранить это различие в оптическом пути, не позволяющее работать с полосами в белом свете, между D и M_2 ставится компенсирующая пластинка C из такого же материала и с такой же толщиной, как у пластинки D, и параллельная ей.

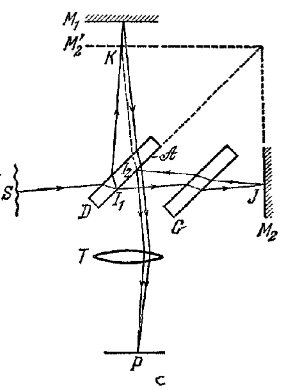


Рис. 7.39. Интерферометр Майкельсона.

Предположим, что M_2'—изображение M_2 поверхностью, делящей пучок. Оптическая длина пути между S и точкой P вдоль луча SI_1JI_2P, отразившегося от M_2, равна оптической длине пути между S и P вдоль луча SI_1KI_2P, отразившегося от мнимой поверхности M_2'. Следовательно, можно считать, что интерференционная картина, наблюдаемая в зрительную трубу, возникает из-за воздушного слоя, ограниченного действительной отражающей поверхностью M_1 и мнимой отражающей поверхностью M_2'. При этом мы должны связать с отражением от M_2 изменение фазы φ, равное разности между изменениями фаз при внешнем и внутреннем отражениях поверхностью $\mathcal{A}$. Величина φ зависит от природы полуотражающего слоя $\mathcal{A}$.

При параллельных зеркалах M_1 и M_2' полосы в квазимонохроматическом свете имеют вид окружностей и локализованы в бесконечности. Они отличаются от полос равного наклона, рассмотренных в п. 7.5.1, только тем, что здесь нет многократных отражений, и, значит, распределение интенсивности находится в полном согласии с (7.2.15). Если M_1 приближается к M_2', полосы стягиваются по направлению к центру, а угловой масштаб картины возрастает до тех пор, пока M_1 не совпадет с M_2'; при этом освещенность поля зрения становится равномерной и ее уровень зависит от φ. Тогда говорят, что зеркала M_1 и M_2' находятся в *оптическом контакте*. Когда M_1 и M_2' близки и образуют клин с небольшим углом, полосы локализуются либо на поверхности клина, либо вблизи нее. При достаточно малом расстоянии между M_1 и M_2' они являются полосами равной толщины и, следовательно, имеют вид эквидистантных прямых линий, параллельных ребру клина. При увеличении расстояния диапазон углов падения, соответствующих каждой точке поля зрения, и вариации среднего угла падения по полю зрения перестают быть незначительными, видность полос уменьшается, и они искривляются, обращаясь выпуклостью в сторону ребра клина.

Независимо от того, параллельны ли M_1 и M_2' или наклонены друг к другу, изменение оптической длины пути в любом плече интерферометра на $\Delta m \cdot \lambda_0$ вызывает смещение картины на Δm порядков. Смещение можно оценить

визуально с точностью до 1/20 порядка, но в определенных условиях методом Кеннеди [23] удается обнаружить смещения до 1/1000 порядка [24].

При расстоянии между M_1 и M_2', не превышающем нескольких длин волн, полосы наблюдаются в белом свете. Они служат для распознавания нулевой полосы в монохроматической интерференционной картине. Если $|\varphi|=\pi$, центральная полоса картины в белом свете оказывается темной и соответствует пересечению M_1 и M_2'; она находится в том же месте, где и полоса в монохроматическом свете с $|m|=1/2$. В общем случае ахроматическая полоса не совпадает со светлой или темной полосой монохроматической картины, но это не вносит дополнительных трудностей, если переход от картины в белом свете к картине в монохроматическом свете происходит постепенно.

В настоящее время интерферометр Майкельсона устарел. Исключение представляет только его модификация, в которой используется освещение параллельным пучком (см. ниже, п. 7.5.5). Несмотря на это, интерферометр такого типа широко известен, так как с его помощью Майкельсон провел три важнейших эксперимента — опыт Майкельсона — Морлея по увлечению эфира [25], первое систематическое изучение тонкой структуры спектральных линий, первое прямое сравнение длин волн спектральных линий с эталонным метром [26]. В нашей книге мы не затронем ни первой, ни третьей из названных работ, так как, во-первых, здесь мы занимаемся только вопросами оптики неподвижных сред, и, во-вторых, после Майкельсона более точные измерения длин волн были выполнены другими методами (см. § 7.7). Метод Майкельсона, касающийся анализа спектральных линий, также был заменен позднее более прямыми методами, но вследствие его значительного теоретического интереса и вследствие его связи с теорией частичной когерентности мы позже рассмотрим его подробно (см. п. 7.5.8).

7.5.5. Интерферометр Тваймана — Грина и другие аналогичные приборы. Если интерферометр Майкельсона освещается точечным источником S квазимонохроматического света, находящимся в фокусе хорошо скорректированной линзы L_1, а выходящий из интерферометра свет собирается второй такой же хорошей линзой L_2, то такой прибор становится эквивалентным интерферометру Физо, но в отличие от последнего здесь пути световых пучков полностью разделены (рис. 7.40). Пусть W_1 — плоский волновой фронт пучка, отраженного от M_1, W_2 — соответствующий плоский волновой фронт пучка, отраженного от M_2, и W_1' — виртуальный плоский волновой фронт, распространяющийся от M_2, который должен был бы выйти из делителя пучка совпадающим и синфазным с W_1. Оптическая разность хода между выходящими после делителя лучами, имеющими виртуальное пересечение в точке P на W_2, равна

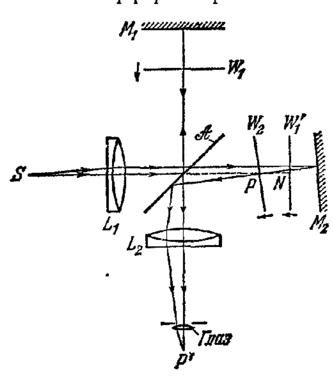


Рис. 7.40. Интерферометр Майкельсона, освещенный параллельным пучком света.

$$\Delta\mathcal{S}=nh, \qquad (74)$$

где $h=PN$ — расстояние по нормали от W_1' до P, а n — показатель преломления среды между W_1' и W_2. Соответствующая разность фаз равна

$$\delta=\frac{2\pi}{\lambda_0}nh. \qquad (75)$$

В соответствии с последним соотношением и (7.2.16) находим, что глаз, помещенный в фокальную плоскость линзы L_2 и сфокусированный на W_2 (в случае надобности с помощью вспомогательной линзы), увидит в P светлую полосу,

если

$$nh = m\lambda_0, \quad |m| = 0,\ 1,\ 2, \ldots, \tag{76а}$$

или темную полосу, если

$$nh = m\lambda_0, \quad |m| = {}^1/_2,\ {}^3/_2,\ {}^5/_2, \ldots \tag{76б}$$

Таким образом, полосы, вообще говоря, имеют вид прямых линий, параллельных ребру клина, образованного W_1' и W_2. Если путем соответствующей ориентации M_1 сделать параллельными фронты W_1' и W_2, то поле зрения будет освещено равномерно. При точечном источнике полосы не локализованы, но практически из-за недостатка света размеры источника нельзя сделать очень малыми. Так как пути выходящих лучей соответствуют отражению от клина, образованного M_2 и M_1', где M_1'— мнимое изображение M_1 в делителе пучка, то полосы с протяженным источником кажутся локализованными вблизи клина, так же как и полосы Физо (см. стр. 276), и допустимые размеры источника наиболее велики, если M_1' и M_2 совпадают. Вместе с тем вследствие неполной монохроматичности источника полосы будут наблюдаться только в случае приблизительного равенства оптических длин путей обоих пучков. Важно отметить, что условия равенства оптических длин путей и совпадения M_1' и M_2, вообще говоря, не могут выполняться одновременно, если установка асимметрична относительно $\mathcal{A}$.

Такое видоизменение интерферометра Майкельсона было предложено Тwайманом и Грином [27] для проверки качества оптических деталей. Последние помещают в пучок, идущий к M_2, и, если они хороши, то отраженный волновой фронт W_2 остается плоским. Так как, согласно (76), мы вправе считать,

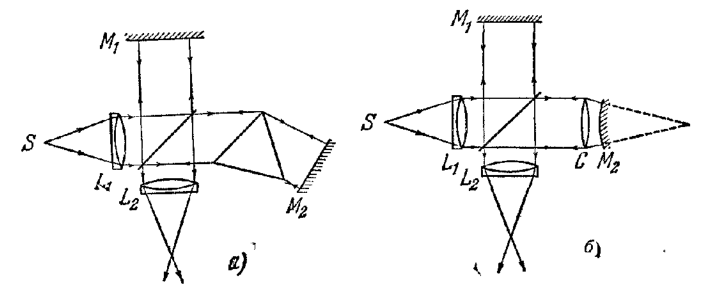


Рис. 7.41. Интерферометр Тваймана — Грина.

а — устройство для испытания призм; б — устройство для испытания фотографических объективов.

что светлые полосы соответствуют контурам W_2 с интервалами λ_0, определяемым плоскостями, параллельными W_1', то (принимая для воздуха $n = 1$) можно обнаружить искажения W_2, получившиеся в результате двойного прохождения света сквозь испытуемую деталь. «Знак» искажения определяется направлением смещения полос при увеличении расстояния между M_1 и делителем пучка. Схема прибора для испытания призм в минимуме отклонения показана на рис. 7.41, *а* [28]. Наблюдаются полосы, совпадающие с одной из поверхностей призмы, а по их положению можно указать те участки призмы, которые должны подвергнуться ретуши. Таким путем можно скомпенсировать внутренние неоднородности материала призмы. На рис. 7.41, *б* показана схема приспособления для испытания фотографических объективов [29]. M_2— выпуклое сферическое зеркало с центром кривизны в фокусе испытываемого объектива C. Объектив может вращаться вокруг линии, перпендикулярной его оси, что позволяет проводить испытания при различных наклонах объектива. Механические связи обеспечивают неизменность положения центра кривизны зеркала M_2

в фокальной плоскости объектива при его вращении. Однако существует некоторая неопределенность в определении по интерференционной картине дефектных мест в объективе, так как в случае несовершенства объектива прямой и

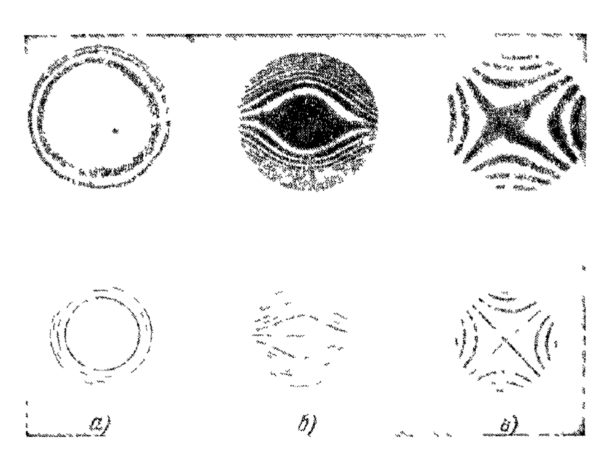


Рис. 7.42. Интерференционные картины Тваймана — Грина, получающиеся с линзой и показывающие наличие первичных аберраций [30].

а — сферическая аберрация, $\Phi_{макс}=0.4\lambda_0$, фокальная параксиальная плоскость; б — кома, $\Phi_{макс}=2\lambda_0$, фокальная параксиальная плоскость, в — астигматизм, $\Phi_{макс}=2,1\lambda_0$, центральная плоскость $\Phi_{макс}$ — максимальная волновая аберрация в выходном зрачке. Вверху — наблюдаемые картины, внизу — рассчитанные.

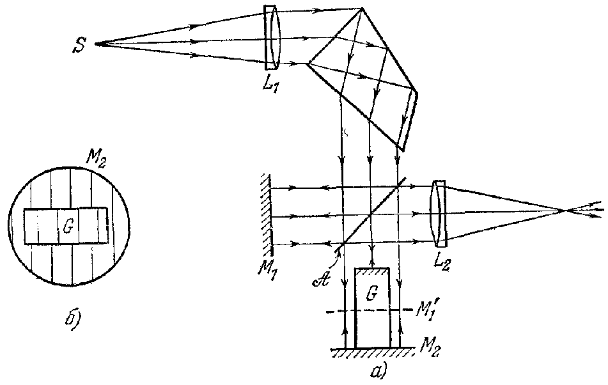


Рис. 7.43. Интерферометр Кюстерса.

а — схема установки, б — поле зрения

обратный пути света через него не совпадают. Можно уменьшить такую неопределенность, выбирая радиус кривизны зеркала M_2 возможно большим и используя для наблюдения виртуальной картины полос, возникающих на его поверхности, соответствующую оптическую систему. На рис. 7.42 приведены фотографии таких картин и соответствующие рассчитанные картины.

Принцип устройства Тваймана — Грина использован в *интерферометре Кюстерса*, предназначенном для измерения длин концевых калибров [31] (рис. 7.43, *а*). Калибр *G* плотно прижимают к зеркалу M_2; зеркало M_1 распо-

лагают в таком месте, чтобы эквивалентная виртуальная отражающая поверхность M_1' находилась примерно на половине пути от M_2 до верхней поверхности G и была наклонена относительно G и M_2 на некоторый небольшой угол. В квазимонохроматическом свете будут видны параллельные эквидистантные полосы на M_2 и G, перпендикулярные к одному из краев G (рис. 7.43, *б*). Длина калибра G определяется методом дробных частей порядка, как и в случае уже описанного интерферометра английской Национальной физической лаборатории (см. п. 7.5.2). Совершенно очевидно, что между этими двумя приборами существует большое сходство *).

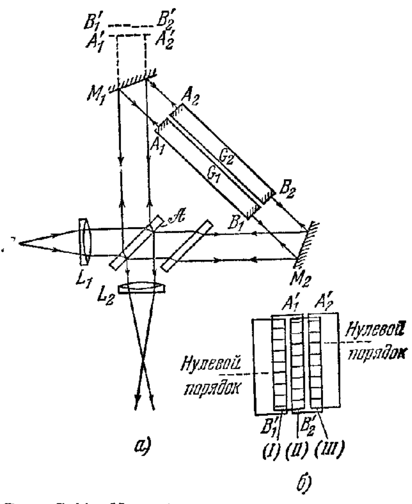


Рис. 7.44. Интерферометр Дауэлла для измерения концевых мер.
а — схема установки, *б* — поле зрения.

Можно упомянуть еще один подобный прибор для измерения концевых мер — *интерферометр Дауэлла* [32] (рис. 7.44, *а*), позволяющий сравнивать длины двух калибров без необходимости тщательного прижатия. Пусть A_1' и A_2' — мнимые изображения в зеркале M_1 поверхностей A_1 и A_2 калибров G_1 и G_2, а B_1' и B_2' — мнимые изображения в M_2 и в делителе пучка поверхностей калибров B_1 и B_2. Расположение деталей прибора таково, что мнимые изображения A_1', A_2' и B_1', B_2' отражающих поверхностей калибров представляются налагающимися друг на друга (рис. 7.44, *б*), и можно считать, что интерференционные полосы возникли в клинообразном воздушном слое между поверхностями A_1', A_2', B_1' и B_2'. Соответствующей установкой G_1 получаем картину горизонтальных полос в монохроматическом свете *I* в области перекрытия A_1' и B_1', и полоса нулевого порядка, которую можно отождествить в белом свете, перемещается в центр поля. Затем калибр G_2 устанавливают таким образом, чтобы A_2' было в одной плоскости с A_1'. Тогда, поскольку поверхность B_2' параллельна B_1', образуются сходные картины горизонтальных полос *II* и *III* в областях перекрытия B_2' с A_1' и A_2' с совпадающими нулевыми порядками. При таких условиях расстояние d между B_2' и B_1', равное разнице в длинах калибров, определяется из соотношения $d = \lambda_0 \Delta m/2n$, где Δm — смещение порядков в картинах *I* и *II*. С этим прибором, применяя метод дробных частей порядка, можно измерять и длины одиночных калибров.

7.5.6. Полосы, получающиеся с двумя одинаковыми пластинами; интерферометр Жамена и интерференционный микроскоп. Пусть свет от квазимонохроматического точечного источника S падает на две прозрачные плоскопараллельные пластинки (толщиной h каждая и с показателем преломления n'), расположенные друг за другом и образующие между собой небольшой угол α (рис. 7.45). Если пренебречь лучами, испытавшими более двух отражений от

*) У обоих приборов существует предел в увеличении расстояний между отражающими поверхностями, до которого полосы еще удовлетворительно видны. Этот предел зависит от углового размера источника (см. п. 7.5.3) и от степени его немонохроматичности (см. п. 7.5.8). Так как в устройстве Кюстерса используется вспомогательная отражающая поверхность M_1', лежащая посередине калибра, то с его помощью при данном источнике можно измерять калибры почти в два раза длиннее, чем с прибором английской Национальной физической лаборатории.

поверхностей пластинок, то останутся две группы параллельных лучей, образовавшихся из падающего луча SA и прошедших сквозь пластинки. Первая группа состоит из прошедшего луча и лучей, отраженных обеими поверхностями каждой пластинки. Вторая группа — $SABDH$, $SABCFI$, $SABDEJ$ и $SABCFGK$ — из лучей, отраженных от одной поверхности каждой пластинки; они три раза проходят промежуток между пластинками и образуют с лучами первой группы угол 2α.

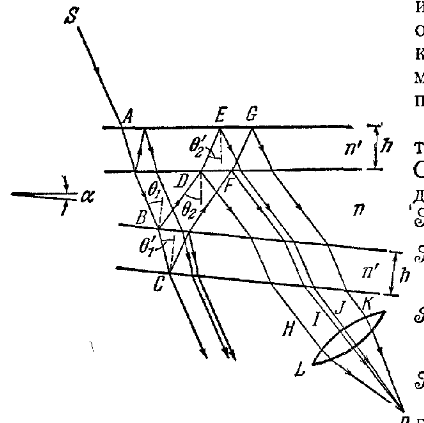


Рис. 7.45. Две одинаковые плоскопараллельные пластинки; возникновение полос, локализованных в бесконечности.

Лучи второй группы собираются в точке P в фокальной плоскости линзы L. Обозначив оптические длины путей от S до P вдоль этих лучей через $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ и $\mathcal{S}_4$, получим, пользуясь (7.5.5),

$$\mathcal{S}_2-\mathcal{S}_1=\Delta\mathcal{S}_{21}=2n'h\cos\theta_1', \tag{77а}$$

$$\mathcal{S}_3-\mathcal{S}_2=\Delta\mathcal{S}_{32}=2n'h(\cos\theta_2'-\cos\theta_1'), \tag{77б}$$

$$\mathcal{S}_4-\mathcal{S}_3=\Delta\mathcal{S}_{43}=2n'h\cos\theta_1', \tag{77в}$$

где θ_1' и θ_2' — углы преломления в точке B второй пластинки и в точке D первой пластинки соответственно; так как для данной точки P углы θ_1' и θ_2' зависят только от ориентации пластинок, то оптические разности хода $\Delta\mathcal{S}$ не зависят от положения S, и интерференционная картина от протяженного источника образуется в фокальной плоскости линзы L. Если диапазон углов падения не слишком велик, эта картина не перекрывается картиной, создаваемой лучами первой группы. Далее, так как $\cos\theta_1\approx\cos\theta_2'$, а $\Delta\mathcal{S}_{21}$ и $\Delta\mathcal{S}_{43}$ велики по сравнению с $\Delta\mathcal{S}_{32}$, при достаточно большом h мы можем использовать источник, для которого условие четкости полос (7.3.15) справедливо по отношению к $\Delta\mathcal{S}_{32}$, но не к $\Delta\mathcal{S}_{21}$ и $\Delta\mathcal{S}_{43}$. В этих условиях мы можем считать, что картина полос связана только с такими лучами как $SABCFI$ и $SABDEJ$. Остальной свет создает почти равномерный фон, просто уменьшающий видность полос.

Из (77б) находим, что разность фаз в точке P, соответствующая $\Delta\mathcal{S}_{32}$, равна

$$\delta=\frac{4\pi}{\lambda_0}n'h(\cos\theta_2'-\cos\theta_1'), \tag{78}$$

т. е. на основании закона преломления

$$\delta=\frac{4\pi h}{\lambda_0}\left(\sqrt{n'^2-n^2\sin^2\theta_2}-\sqrt{n'^2-n^2\sin^2\theta_1}\right), \tag{79}$$

где θ_1 и θ_2 — углы падения соответственно в точках B и D, а n — показатель преломления среды, окружающей пластинки. Для того чтоб определить форму полос, рассмотрим прямоугольные оси координат OX, OY и OZ с началом в точке O во второй узловой точке линзы L, причем ось OZ параллельна ребру клина, образованного пластинами (рис. 7.46). Пусть N_1 в плоскости XOY — фокальная точка для света, отраженного нормально в B ($\theta_1=0$), а N_2 (также в плоскости XOY) — фокальная точка для лучей, отраженных нормально в D ($\theta_2=0$) Пусть далее OX — биссектриса угла N_1ON_2. Тогда, если клин «раскрыт» в направлении OY, направляющие косинусы прямых ON_1 и ON_2 равны соответственно $\{\cos(\alpha/2), \sin(\alpha/2), 0\}$, $\{\cos(\alpha/2), -\sin(\alpha/2), 0\}$. Пусть OP составляет угол ψ с плоскостью XOZ, а проекция OP на нее составляет угол φ с OX. Тогда

направляющие косинусы прямой OP равны ($\cos\psi\cos\varphi$, $\sin\psi$, $\cos\psi\sin\varphi$), и так как $PON_1 = \theta_1$, $PON_2 = \theta_2$, получаем

$$\cos\theta_1 = \cos\frac{\alpha}{2}\cos\psi\cos\varphi + \sin\frac{\alpha}{2}\sin\psi, \tag{80а}$$

$$\cos\theta_2 = \cos\frac{\alpha}{2}\cos\psi\cos\varphi - \sin\frac{\alpha}{2}\sin\psi. \tag{80б}$$

При малом α мы можем пренебречь членами, содержащими вторую и более высокие степени α в разложении $\cos\left(\frac{\alpha}{2}\right)$, $\sin\left(\frac{\alpha}{2}\right)$, и из (80) находим

$$\cos^2\theta_1 = \cos^2\psi\cos^2\varphi + \alpha\sin\psi\cos\psi\cos\varphi, \tag{81а}$$

$$\cos^2\theta_2 = \cos^2\psi\cos^2\varphi - \alpha\sin\psi\cos\psi\cos\varphi. \tag{81б}$$

Тогда, учитывая (79) и последние соотношения, имеем

$$\delta = \frac{4\pi h}{\lambda_0}\left\{\sqrt{n'^2 - n^2(1-\cos^2\psi\cos^2\varphi + \alpha\sin\psi\cos\psi\cos\varphi)} - \right.$$

$$\left. - \sqrt{n'^2 - n^2(1-\cos^2\psi\cos^2\varphi - \alpha\sin\psi\cos\psi\cos\varphi)}\right\} \tag{82}$$

и после разложения по степеням α, вновь отбрасывая члены с α в степени выше первой, получим

$$\delta = -\frac{4\pi h}{\lambda_0}\frac{n^2\alpha\sin\psi\cos\psi\cos\varphi}{\sqrt{n'^2 - n^2(1-\cos^2\psi\cos^2\varphi)}}. \tag{83}$$

Поэтому, согласно (83) и (7.2.16), светлые полосы в точке P будут при

$$\frac{\sin\psi\cos\psi\cos\varphi}{\sqrt{n'^2 - n^2(1-\cos^2\psi\cos^2\varphi)}} = \frac{m\lambda_0}{2n^2h\alpha},$$

$$|m| = 0,\ 1,\ 2,\ \ldots, \tag{84а}$$

а темные полосы — при

$$\frac{\sin\psi\cos\psi\cos\varphi}{\sqrt{n'^2 - n^2(1-\cos^2\psi\cos^2\varphi)}} = \frac{m\lambda_0}{2n^2h\alpha},$$

$$|m| = 1/2,\ 3/2,\ 5/2,\ \ldots \tag{84б}$$

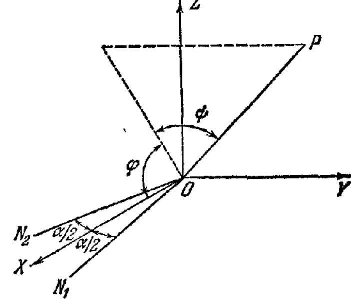


Рис. 7.46. К исследованию локализованных в бесконечности полос, получающихся с двумя плоскопараллельными пластинками, наклоненными друг к другу.

Если точка P находится близко к плоскости XOZ, так что ψ мало, порядок интерференции m низок, и в этой области полосы можно наблюдать в белом свете. Центральная светлая полоса ($m = 0$) лежит в плоскости XOZ и образуется лучами, для которых $\theta_1 = \theta_2$.

В особом случае наблюдения в направлении, близком к нормальному падению, ψ и φ малы, и, отбрасывая члены со второй и более высокой степенью ψ и φ, мы сведем (84а) к

$$\psi = \frac{mn'\lambda_0}{2n^2h\alpha}, \quad |m| = 0,\ 1,\ 2,\ \ldots \tag{85}$$

Таким образом, полосы в квазимонохроматическом свете имеют вид эквидистантных прямых линий, параллельных ребру клина между пластинками. Угловое расстояние между соседними светлыми полосами пропорционально показателю преломления пластинок и обратно пропорционально их толщине и углу между ними. Эти полосы впервые наблюдались Брюстером и называются *полосами Брюстера*.

Подобные полосы используются в *интерферометре Жамена* [33], широко применявшемся одно время для измерений показателя преломления газов. Теперь этот прибор заменен интерферометром Рэлея (см. п. 7.3.5). Основные

части прибора — две стеклянные плоскопараллельные пластины одинаковой толщины с одинаковым показателем преломления. Поверхности пластин M_1 и M_2 покрыты непрозрачным зеркальным слоем серебра; их расположение показано на рис. 7.47. Пучок света от протяженного источника падает под углом, близким к 45°, на одну из пластин. В результате отражений и преломлений возникают два пучка: один, отраженный от передней поверхности первой пластины и задней поверхности второй, и другой, отраженный от задней поверхности первой пластины и передней поверхности второй. Эти пучки образуют интерференционную картину в фокальной плоскости зрительной трубы T. Толщина пластин выбирается такой, чтобы при подходящем диафрагмировании пучки между пластинами полностью разделялись и чтобы каждый пучок проходил через одну из газовых кювет G_1, G_2 и один из компенсаторов C_1, C_2, подобных описанным в п. 7.3.5. В работе пластины слегка наклоняют, так что они образуют друг с другом небольшой клин с ребром, параллельным плоскости рис. 7.47, которую мы будем считать горизонтальной; поверхности пластин почти вертикальны, и, следовательно, пользуясь теми же обозначениями, что и на рис. 7.46, мы можем считать плоскость XOZ почти горизонтальной. Область наблюдения в этом случае соответствует малому ψ, и положение светлых полос определяется, согласно (84а), соотношением

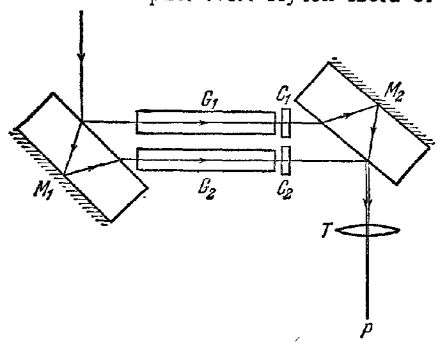


Рис. 7.47. Интерферометр Жамена.

$$\frac{\cos\varphi}{\sqrt{n'^2-n^2\sin^2\varphi}}\,\psi=\frac{m\lambda_0}{2n^2h\alpha},\qquad |m|=0,\ 1,\ 2,\ldots \tag{86}$$

Так как $\varphi\approx 45°$, то при угловом поле небольшого размера изменения члена с φ пренебрежимо малы. Следовательно, полосы соответствуют точкам с одинаковым ψ, т. е. они горизонтальны и эквидистантны. Эти полосы относятся к низким порядкам интерференции; соответствующей установкой полосу нулевого порядка ($\psi=0$), отождествленную в белом свете, можно сместить в центр поля зрения *).

Метод измерения с интерферометром такого типа подобен уже описанному методу работы с интерферометром Рэлея. Разница заключается только в том, что здесь отсутствует вторая система полос, служащая неподвижными реперами, и установка ведется по кресту нитей в зрительной трубе. Вследствие этого прибор более чувствителен к нарушениям в оптической системе, чем интерферометр Рэлея, и точность получаемых с его помощью измерений меньше.

Сиркс [34] и позднее Прингсгейм [35] предложили для измерения показателя преломления малых объектов видоизмененный интерферометр Жамена со слегка клиновидными, а не плоскопараллельными пластинами. Пластины устанавливают таким образом, чтобы ребра клиньев были антипараллельны,

*) Если пластины наклонены так, что ребро образованного ими клина вертикально, то плоскость XOY горизонтальна, и область наблюдения соответствует малому φ, а положение максимумов интенсивности определяется соотношением (84а), т. е.

$$\frac{\cos\psi\sin\psi}{\sqrt{n'^2-n^2\sin\psi}}=\frac{m\lambda_0}{2n^2h\alpha},\qquad |m|=0,1,2\ldots$$

Полосы соответствуют точкам с одинаковым ψ и в нашем случае будут вертикальны. Однако при $\psi\sim 45°$ порядок интерференции значительно отличается от нулевого и с таким расположением пластин полосы в белом свете не наблюдаются.

а внутренние несеребренные поверхности — почти параллельны (рис. 7.48), и освещают параллельным пучком света. Падающему лучу SA, лежащему в главном сечении клиньев, соответствуют два луча: $SABCG$ и $SADEF$, которые после выхода из второй пластины пересекаются в мнимой точке P, лежащей сзади нее. С квазимонохроматическим не очень большим источником интерференционные полосы, кажущиеся расположенными вблизи P, можно наблюдать в микроскоп M. В этой области полосы направлены под прямым углом к плоскости, определяемой обоими выходящими из пластин лучами, т. е. параллельны ребрам клиньев. Исследуемый объект O помещают между пластинами на пути луча CG. Изображение P' точки P передней поверхностью второй пластины также лежат на луче CG, а местоположение этого изображения зависит от наклона пластин. По Прингсгейму [35] точка P' находится приблизительно на половине пути между пластинами, если их внутренние поверхности параллельны, но незначительный поворот любой пластины вокруг оси, параллельной ребру клина, вызывает заметное смещение P'. Таким образом, точку P' можно совместить с O, и в поле зрения микроскопа изображение объекта и система интерференционных полос наложатся друг на друга. Изменение порядка интерференции в точке P', вызванное введением объекта O, определяется компенсатором. Отсюда, зная толщину объекта в P', можно определить и его показатель преломления.

Сравнительно недавно Дайсон [36] разработал аналогичную комбинацию

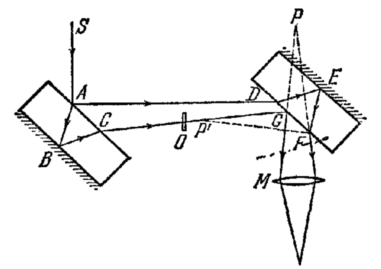


Рис. 7.48. Интерферометр Сиркса — Прингсгейма.

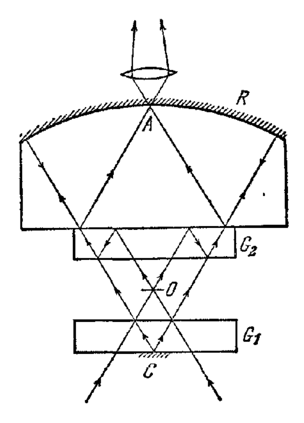


Рис. 7.49. Интерференционный микроскоп Дайсона.

микроскопа и интерферометра (рис. 7.49). Исследуемый предмет, находящийся на предметном стекле, помещают в O между двумя одинаковыми стеклянными пластинами G_1 и G_2. Пластины слегка клиновидны и их ребра антипараллельны. Верхняя поверхность нижней пластины G_1 покрыта полупрозрачным слоем серебра, а ее нижняя поверхность на небольшом центральном участке C, немного большем поля зрения микроскопа, покрыта непрозрачным слоем серебра. Верхняя пластина G_2 покрыта частично прозрачным слоем серебра с обеих сторон. Пространство между пластинами (в том числе и область, где расположен исследуемый предмет) заполняется веществом с таким же показателем преломления, как и у стекла. Свет проходит через конденсор микроскопа, фокусирующий изображение источника на плоскость O. Часть этого света — предметный пучок — проходит через O и выходит из G_2 после отражения от его верхней и нижней поверхностей. Другая его часть — опорный пучок — отражается от верхней поверхности G_1 и сходится на C, где он отражается, и, проходя вне предмета O, который находится в тени C, прямо проходит через G_2. Стеклянный блок со сферической верхней поверхностью R, покрытой (за исключе-

нием небольшого участка A на оси в верхней части R) непрозрачным слоем серебра, склеен с G_2. Он рассчитан так, чтобы предметный и опорный пучки после отражения на R и на верхней поверхности G_2 сходились в A. Опорный пучок образует действительное изображение σ_1 источника вблизи A, предметный пучок — наложенное друг на друга действительное изображение σ_2 источника и действительное изображение предметной плоскости. Последнее рассматривают в обычный микроскоп *).

Соответствующие точки на σ_1 и σ_2 являются изображениями одних и тех же точек источника и поэтому ведут себя как взаимно когерентные вторичные источники. Прибор юстируют таким образом, чтобы эти точки совпали. В отсутствие предмета оптические длины путей предметного пучка и пучка сравнения в любой точке P на П равны и, пренебрегая небольшой разностью фаз, вызванной различием в отражениях на серебреных поверхностях, находим, что порядок интерференции на всем изображении П будет нулевым. Это можно проверить в белом свете. В присутствии предмета в O оптическая длина пути предметного пучка до P увеличивается приблизительно на $(n'-n)l$, где n'— показатель преломления предмета, l — его толщина в точке P', сопряженной с P, n — показатель преломления окружающей среды. Вариации n' и l по предмету влекут за собой, таким образом, вариации интенсивности по П в квазимонохроматическом свете или изменение цвета в белом свете. При движении нижней пластинки G_1 нормально к ребру клина и оптической оси длина оптического пути пучка сравнения изменяется, и таким способом можно компенсировать изменение длины оптического пути предметного пучка. Проградуировав такое перемещение G_1 в монохроматическом свете и зная n и l, можно определить $(n'-n)\,l$, а значит, и n'.

Однако в реальных условиях полностью совместить σ_1 и σ_2 никогда не удается. К тому же присутствие преломляющего предмета и использование пластины G_1 в качестве компенсатора вызывает взаимное смещение σ_1 и σ_2, параллельное оптической оси. Удовлетворительное описание интерференционных явлений в области σ_1 и σ_2 при неточном совпадении соответствующих точек этих изображений нельзя сделать в рамках элементарного изложения, приведенного в настоящей главе. Его можно изящно рассмотреть с помощью теории частичной когерентности (см. гл. 10).

7.5.7. Интерферометр Маха — Цендера; интерферометр Бейтса со смещенным волновым фронтом. В интерферометре Жамена (см. п. 7.5.6) передние поверхности пластин, выполняющие роль делителей световых пучков, и задние поверхности, служащие плоскими зеркалами, нельзя установить независимо и, следовательно, расстояние между пучками определяется толщиной пластин. Значительно большей гибкостью обладает прибор, в котором делители пучков и зеркала представляют собой независимые элементы, а пучки можно широко развести. На этом принципе основано устройство *интерферометра Маха — Цендера* [37], применяемого для измерений изменений показателя преломления, а следовательно, и плотности потоков сжимаемого газа.

Схема такого прибора показана на рис. 7.50. Свет от источника S, находящегося в фокальной плоскости хорошо коррегированной линзы L_1, делится на полуотражающей поверхности $\mathcal{A}_1$ плоскопараллельной стеклянной пластинки D_1 на два пучка. Отразившись от плоских зеркал M_1 и M_2, они вновь соединяются на полуотражающей поверхности $\mathcal{A}_2$ второй (идентичной D_1) плоскопараллельной пластинке D_2 и выходят через вторую тоже хорошо коррегированную линзу L_2. Все четыре отражающие поверхности обычно устанавливают почти параллельно друг другу и так, чтобы их центры находились в вершинах параллелограмма. Предположим теперь, что S — точечный источник квазимонохроматического света. Пусть W_1— плоский волновой фронт пучка

*) Для всех остальных пучков света, кроме упомянутых, несеребреный участок A полностью экранируется C, и свет этих пучков не может попасть в микроскоп.

между M_1 и D_2, W_2— соответствующий плоский волновой фронт пучка между M_2 и D_2 и W_1— виртуальный плоский волновой фронт между M_2 и D_2, который, выходя из D_2, совпадает с W_1 и синфазен ему. В точке P на волновом фронте W_2 виртуальная разность фаз между выходящими пучками равна

$$\delta = \frac{2\pi}{\lambda_0} nh, \tag{87}$$

где $h = PN$ — расстояние по нормали от P до W_1', а n — показатель преломления среды между W_2 и W_1'. Согласно (7.2.16) в сопряженной с P точке P' в выходящих пучках светлая полоса находится при

$$nh = m\lambda_0, \qquad |m| = 0,\ 1,\ 2, \dots \tag{88а}$$

темная полоса — при

$$nh = m\lambda_0, \qquad |m| = \frac{1}{2}, \frac{3}{2}, \frac{5}{3}, \dots \tag{88б}$$

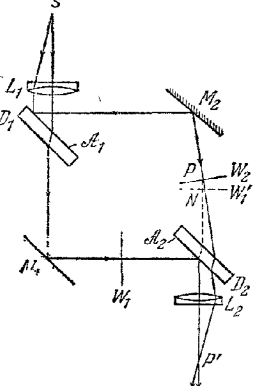


Рис. 7.50. Интерферометр Маха — Цендера.

Если W_1' и W_2 параллельны, интенсивность во всех точках P одинакова, и при этих условиях протяженный источник даст полосы в бесконечности (т. е. в фокальной плоскости L_2), так же как и интерферометр Жамена. Однако в общем случае W_1' и W_2 составляют некоторый угол, и полосы имеют вид прямых, параллельных линии пересечения W_1' и W_2. Именно такая установка применяется при исследовании газовых потоков, где обычно ощущается недостаток света, и поэтому желательно применять наибольшие из возможных источников, обеспечивающих еще достаточную видность. Как мы видели в п. 7.5.3, полосы при этом локализуются в области, где пересекающиеся лучи имеют наименьшее угловое расхождение при выходе из S. Положение области локализации изменяется при повороте элементов системы, которые служат для изменения взаимного наклона W_1' и W_2. Например, если отражающие поверхности первоначально были параллельны и если для простоты мы рассмотрим случай поворота вокруг оси, перпендикулярной к плоскости, в которой лежат все центры, то при повороте M_2 виртуальный район локализации находится вблизи M_2 (рис. 7.51, *а*), а при повороте и M_2 и D_2— между M_2 и D_2 (рис. 7.51, *б*). Эта особенность отличает полосы, возникающие в клине интерферометра Маха — Цендера, от полос, появляющихся в интерферометре Майкельсона при освещении его параллельным пучком света (см. п. 7.5.5) с виртуальной областью локализации близ зеркал.

В техническом варианте прибора в одном плече интерферометра помещается камера C_1 для исследования газового потока — обычно рабочая секция аэродинамической трубы или трубы для изучения ударных волн. В другом его плече располагается компенсационная камера C_2, позволяющая получить полосы порядков, близких к нулю, желаемой ориентации, с удобным расстоянием и с виртуальной локализацией вблизи выбранной плоскости в C_1, нормальной к направлению падающего света. Эта плоскость изображается на фотографической пластинке посредством линзы L_2 и хорошо коррегированного фотографического объектива. Фотосъемка картины интерференционных полос производится как в присутствии газового потока, так и без него, и в выбранной точке P' плоскости изображения измеряется смещение порядка Δm в двух картинах. Если нужно, то используются полосы в белом свете для идентификации соответствующих порядков. Пусть n — показатель преломления невозмущенного газа в C_1, а n'— показатель преломления текущего газа; тогда

$$\Delta m = \frac{1}{\lambda_0} \int (n' - n)\, ds. \tag{89}$$

Здесь интеграл берется вдоль пути лучей, проходящих через C_1 и достигающих P'. Пусть Ox, Oy, Oz — оси прямоугольной системы координат с началом O в выбранной плоскости в камере C_1, причем ось Oz совпадает с направлением падающего света, пусть далее P с координатами $(x, y, 0)$ — точка, сопряженная с P' в отсутствие газового потока в C_1. Тогда, если пренебречь отклонением лучей, возникающим вследствие преломления из-за наличия потока, то (89) можно переписать в виде

$$\Delta m(x, y) = \frac{1}{\lambda_0}\int_0^s \{n'(x, y, z) - n\}\, dz, \tag{90}$$

где s — длина C_1; если поток удовлетворяет определенным условиям симметрии (см., например, [38]), $(n' - n)$ в (90) можно выразить через измеренные величины

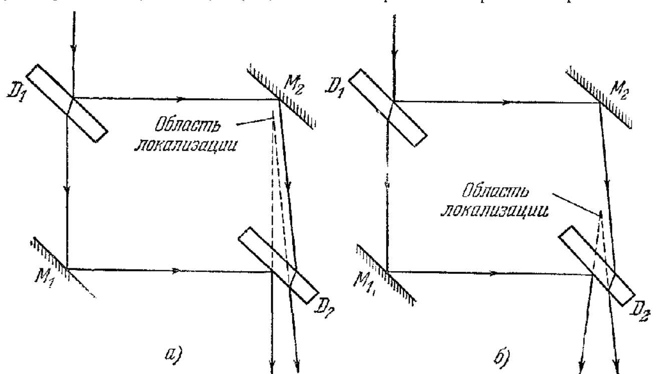


Рис 7 51. К локализации полос в интерферометре Маха — Цендера

Δm. В таком случае можно определить изменение плотности $\Delta\rho$, вызванное потоком, так как оно пропорционально $(n' - n)$.

Видоизмененный Бейтсом [39] интерферометр Маха — Цендера можно применить для измерения асферичности сходящихся волновых фронтов, не прибегая к сравнению со свободным от искажений опорным волновым фронтом, как это необходимо в методе Тваймана — Грина (см. стр. 280). Устройство, показанное на рис. 7.52, *а*, особенно подходит для испытания качества систем с большими апертурами. Подлежащий проверке сходящийся пучок с главной осью OA, направленной горизонтально, делится в D_1 на два сходящихся пучка, дающих изображения S_1 и S_2 достаточно малого квазимонохроматического источника. В начале все четыре отражающие поверхности вертикальны, параллельны и расположены так, что на полуотражающей поверхности D_2 изображения S_1 и S_2 совпадают. Тогда соответствующие падающему волновому фронту W виртуальные волновые фронты W_1 и W_2 с главными осями O_1S_1 и O_2S_2 точно совпадут, и глаз, помещенный после D_2, увидит равномерно освещенное поле зрения. Повернем теперь D_1 и M_1 как целое вокруг оси, параллельной OA; тогда O_1S_1 и O_2S_2 немного сместятся друг относительно друга по вертикали. Это по существу равносильно повороту W_2 относительно W_1 вокруг горизонтальной оси, и мы увидим, что поле зрения пересекается эквидистантными горизонтальными полосами, видимыми в белом свете.

Предположим теперь, что D_2 поворачивается вокруг вертикальной оси, проходящей через S_1 и S_2; тогда O_1S_1 повернется около S_1 в горизонтальной плоскости и W_1 сместится относительно W_2. При идеально сферических W_1 и W_2 на картину полос, наблюдающуюся в области, где перекрывание W_1 и W_2

еще сохранилось, не повлияет смещение полос. При наличии отклонений от сферичности полосы сместятся на расстояние, зависящее от степени асферичности W. Поэтому, если направление оси координат O_2X с началом в O_2 совпадает с направлением смещения, то смещение порядка $\Delta m(x;\ a)$ в точке P с координатой x равно

$$\Delta m(x;\ a)=\frac{1}{\lambda_0}\{\mathcal{S}(x)-\mathcal{S}(x-a)\}, \tag{91}$$

где $\mathcal{S}$ — оптическая разность хода между W_2 и сферой с центром в S_2 и

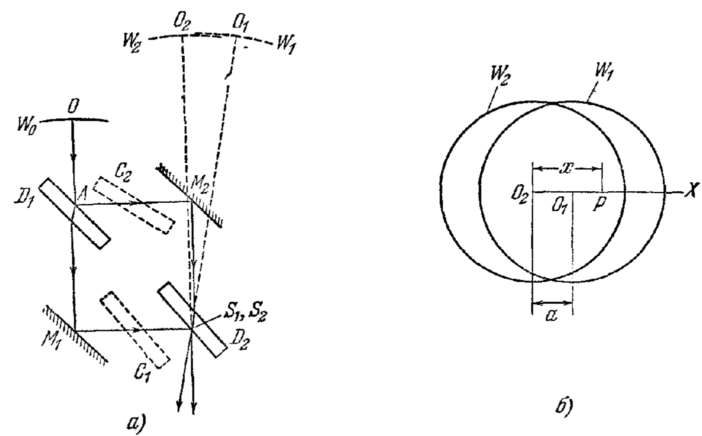


Рис. 7.52. Интерферометр Бейтса со смещенным волновым фронтом.
а — схема установки, б — поле зрения со смещенными волновыми фронтами.

радиусом O_2S_2, а a — величина сдвига (рис. 7.52, б). Так как $\mathcal{S}$ равно нулю в O_2 $(x=0)$, имеем отсюда

$$\Delta m(a;\ a)=\frac{1}{\lambda_0}\mathcal{S}(a) \tag{92а}$$

и аналогично

$$\Delta m(2a;\ a)=\frac{1}{\lambda_0}\{\mathcal{S}(2a)-\mathcal{S}(a)\},$$

так что

$$\Delta m(a;\ a)+\Delta m(2a;\ a)=\frac{1}{\lambda_0}\mathcal{S}(2a). \tag{92б}$$

Подобным же образом нетрудно получить выражения для $\mathcal{S}(3a)$, $\mathcal{S}(4a)$ и т. д., и мы видим, что величину $\mathcal{S}(x)$ можно определить с интервалом a из измерений Δm. В другом случае, когда a не слишком велико, мы получим из (91) соотношение

$$\Delta m(x;\ a)\approx\frac{a}{\lambda_0}\frac{\partial\mathcal{S}(x)}{\partial x}, \tag{93}$$

показывающее, что Δm пропорционально угловой аберрации $\frac{\partial\mathcal{S}}{\partial x}$ луча, выходящего из P. Очевидно, что в отсутствие вращательной симметрии полный волновой фронт в принципе можно исследовать путем изменения направления сдвига.

Когда волновые фронты смещаются друг относительно друга, выходящие лучи, виртуально пересекающиеся в P, проходят делители пучков под разными углами, и если смещение полос зависит только от асферичности, то такое

различие следует скомпенсировать. Для этой цели служат две компенсирующие пластинки, подобные пластинкам, используемым в качестве делителей пучков, и вводимые в оба плеча интерферометра. Одну из них, C_1, укрепляют параллельно D_2 и при сдвиге поворачивают одновременно с ней. Другую, C_2, соединяют с D_2 механической связью и поворачивают в два раза быстрее, чем D_2 в противоположном направлении. Сравнительно недавно Дрю *) разработал упрощенную модель интерферометра, в которой компенсирующие пластинки не нужны.

7.5.8. Длина когерентности; применение двухлучевой интерференции к изучению тонкой структуры спектральных линий. Газ (например, пары кадмия), возбуждаемый в определенных условиях электрическим разрядом, испускает свет, спектр которого состоит из резких ярких линий, разделенных темными промежутками,— так называемый *эмиссионный линейчатый спектр*. Выделим свет одной из этих линий и осветим им, например, интерферометр Майкельсона, установленный так, чтобы образовались кольцевые интерференционные полосы; тогда мы увидим, что полосы становятся отчетливыми, если длины оптических путей обоих интерферирующих пучков примерно одинаковы. При возрастании оптической разности хода видность полос уменьшается (вообще говоря, немонотонно) и в конце концов они исчезают.

Такое исчезновение полос можно объяснить, предположив, что свет спектральной линии недостаточно монохроматичен и состоит из цугов волн конечной длины, *большое количество которых проходит за любой интервал времени, необходимый для наблюдения*. Допустим теперь, что все эти волновые цуги идентичны. Каждый из них, попадая в интерферометр, делится на два цуга равной длины, и если оптическая разность хода в плечах интерферометра больше этой длины, один из цугов пройдет точку наблюдения P раньше, чем другой дойдет до нее. В таком случае в точке P интерференция двух волновых цугов, образовавшихся из *одного*, наблюдаться не будет. В P в любой момент налагаются друг на друга волновые цуги, порожденные *разными* падающими волновыми цугами, и так как они приходят беспорядочно, быстро сменяя друг друга, то их вклад в интерференционный член за относительно большое время, необходимое для наблюдения, в среднем равен нулю.

Можно представить наше объяснение в другом виде, математически более удобном для описания изменений видности полос с разностью хода, применив интеграл Фурье. Пусть $F(t)$ — световое возмущение в некоторой точке в момент времени t, вызванное одиночным волновым цугом, и пусть F равно нулю для $|t| \geqslant t_0$; представим эту функцию в виде интеграла Фурье

$$F(t) = \int_{-\infty}^{\infty} f(\nu) \exp(-2\pi i\nu t)\, d\nu, \tag{94}$$

где, согласно обратному преобразованию Фурье,

$$f(\nu) = \int_{-\infty}^{\infty} F(t) \exp(2\pi i\nu t)\, dt. \tag{95}$$

Если за время наблюдения эту точку минует N таких волновых цугов, то полное световое возмущение можно записать в виде

$$V(t) = \sum_{n=1}^{N} F(t - t_n), \tag{96}$$

где величины t_n— обозначают моменты прихода соответствующего волнового цуга. Средняя интенсивность света за временной интервал $2T$, необходимый

*) См. [40]. Применение интерферометра со смещенным волновым фронтом обсуждается в [40а].

для наблюдения, равна

$$I=\frac{1}{2T}\int_{-T}^{T}|V(t)|^2\,dt\approx\frac{1}{2T}\int_{-\infty}^{\infty}|V(t)|^2\,dt, \tag{97}$$

если T велико по сравнению с t_0 (половиной длительности волнового цуга). Из (94) и (96) следует, что

$$V(t)=\int_{-\infty}^{\infty}v(\nu)\exp(-2\pi i\nu t)\,d\nu, \tag{98}$$

где

$$v(\nu)=f(\nu)\sum_{n=1}^{N}\exp(2\pi i\nu t_n). \tag{99}$$

Следовательно, в соответствии с теоремой Парсеваля имеем

$$\int_{-\infty}^{\infty}|V(t)|^2\,dt=\int_{-\infty}^{\infty}|v(\nu)|^2\,d\nu=\int_{-\infty}^{\infty}|f(\nu)|^2\sum_{n=1}^{N}\sum_{m=1}^{N}\exp[2\pi i\nu(t_n-t_m)]\,d\nu. \tag{100}$$

Мы можем написать

$$\sum_{n=1}^{N}\sum_{m=1}^{N}\exp[2\pi i\nu(t_n-t_m)]=N+\sum_{n\neq m}\exp[2\pi i\nu(t_n-t_m)]=$$

$$=N+2\sum_{n<m}\cos 2\pi\nu(t_n-t_m). \tag{101}$$

Однако, поскольку величины t_n распределяются случайным образом, вероятности того, что члены с косинусом будут положительными или отрицательными, одинаковы. Следовательно, средняя величина двойной суммы в (100) при большом числе таких опытов равна N, и из (97) и (100) получим, что средняя интенсивность равна

$$I=\frac{N}{2T}\int_{-\infty}^{\infty}|f(\nu)|^2\,d\nu, \tag{102}$$

т. е. пропорциональна интегралу от интенсивностей $i(\nu)=|f(\nu)|^2$ (некогерентная суперпозиция) монохроматических компонент, составляющих одиночные волновые цуги *). В интерферометре каждая монохроматическая компонента дает интерференционную картину, описанную в п. 7.5.4, и так как оптическая разность хода увеличивается от нуля, смещения интерференционных картин, обусловленных разными компонентами, увеличиваются вследствие различия в длинах волн. Следовательно, видность полос уменьшается, и они полностью исчезают, когда оптическая разность хода становится достаточно большой.

Эти две возможные интерпретации отсутствия полос при достаточно большой разности хода — в рамках представления о хаотичной последовательности ограниченных волновых цугов или представления о суперпозиции монохроматических компонент, распределенных в некотором частотном диапазоне,— для большинства практических целей эквивалентны, и из предшествующего обсуждения мы можем заключить, что чем длиннее волновые цуги, тем уже частотные диапазоны, в которых фурье-компоненты имеют заметную интенсивность. Проиллюстрируем эту связь на простом примере. Предположим, что длительность всех волновых цугов равна Δt и в течение этого времени $F(t)$ есть простая

*) Строгая формулировка найденного результата дается теоремой Кэмпбелла, хорошо известной в анализе случайных шумов, особенно в связи с дробовым эффектом (флуктуации в интенсивности потока электронов в вакуумных лампах) [41].

периодическая функция с частотой ν_0, т. е.

$$\left.\begin{aligned} F(t) &= f_0 \exp[-2\pi i \nu_0 t], \quad \text{если} \quad |t| \leqslant \Delta t/2, \\ F(t) &= 0, \quad \text{если} \quad |t| > \Delta t/2, \end{aligned}\right\} \tag{103}$$

где f_0 — постоянная. Тогда из (95) и (103) находим

$$f(\nu) = f_0 \int_{-\Delta t/2}^{\Delta t/2} \exp[2\pi i(\nu - \nu_0)t]\,dt = f_0 \Delta t \left[\frac{\sin\{\pi(\nu-\nu_0)\Delta t\}}{\pi(\nu-\nu_0)\Delta t}\right]. \tag{104}$$

График функции $[\sin\{\pi(\nu-\nu_0)\Delta t\}/(\pi(\nu-\nu_0)\Delta t)]^2$, ответственной за распределение интенсивностей фурье-компонент в (103), представлен на рис. 7.53.

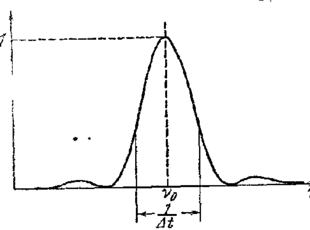


Рис. 7.53. График функции $\left[\frac{\sin\{\pi(\nu-\nu_0)\Delta t\}}{\pi(\nu-\nu_0)\Delta t}\right]^2$.

Частотный интервал $\nu_0 - \Delta\nu/2 \leqslant \nu \leqslant \nu_0 + \Delta\nu/2$, в котором интенсивность можно считать значительной, в известной степени произволен, но так как первый нуль (который появляется, когда аргумент члена с синусом равен π) соответствует $\nu - \nu_0 = \pm 1/\Delta t$, то ясно, что

$$\Delta\nu \approx \frac{1}{\Delta t}. \tag{105}$$

Итак, эффективный частотный диапазон фурье-спектра равен по порядку обратной величине длительности одного волнового цуга.

Данный пример, в котором все волновые цуги идентичны и имеют простую форму, дает лишь идеализированное представление о свете реальных источников. Согласно теории строения атома потеря энергии атомами при излучении вызывает затухание волновых цугов. Далее, атомы находятся в беспорядочном тепловом движении относительно наблюдателя, и поэтому наблюдающиеся спектры искажены эффектом Допплера. Кроме того, излучающие атомы возмущаются соседними, и волновые цуги беспорядочно изменяются. По этим причинам нельзя ожидать, что изучение света реальных источников позволит придать простой смысл терминам «длительность волновых цугов» или «частотный диапазон фурье-спектров». Однако для любого светового возмущения $V(t)$ или его фурье-образа $v(\nu)$ можно определить величины Δt и $\Delta\nu$, которые мы вправе рассматривать соответственно как среднюю длительность волновых цугов, составляющих V, и как эффективный частотный диапазон фурье-спектра; можно показать, что эти средние всегда удовлетворяют соотношению

$$\Delta t \Delta\nu \geqslant \frac{1}{4\pi}. \tag{106}$$

Последнее неравенство, по смыслу аналогичное соотношению неопределенности Гейзенберга в квантовой механике, выводится и обсуждается в п. 10.7.3. Здесь мы отметим только, что в большинстве практически интересных случаев знак неравенства в (106) можно заменить знаком, обозначающим порядок величины.

Промежуток времени Δt, входящий в (106), известен как *время когерентности* света; если $\bar{\lambda}_0$ — средняя длина волны, то длина Δl, определяемая соотношением

$$\Delta l = c\Delta t \sim \frac{c}{\Delta\nu} = \frac{(\bar{\lambda}_0)^2}{\Delta\lambda_0}, \tag{107}$$

известна как *длина когерентности*. Сравнивая (107) и (7.3.15), мы видим, что наше прежнее ограничение, налагаемое на разность хода между квазимонохроматическими пучками означает, что она должна быть мала по сравнению с длиной когерентности света. Если оптическая разность хода того же порядка

или значительно больше длины когерентности, интерференционные эффекты становятся незаметными.

Из предыдущего ясно, что наблюдения за изменением видности полос в зависимости от оптической разности хода в соответствующих интерференционных опытах должны содержать информацию о спектральном распределении интенсивности используемого света. Первые наблюдения такого рода были выполнены Физо [17]. Осветив свой интерферометр (см. п. 7.5.2) желтым светом натриевой лампы, он получил кольца Ньютона и наблюдал за ними при увеличении расстояния между линзой и пластинкой. Физо нашел, что при контакте линзы с пластинкой кольца были четкими, почти исчезли вблизи 490-го кольца и снова приобретали приблизительно первоначальную четкость около 980-го кольца. Он смог проследить периодическое изменение видности полос в 52 периодах из 980 колец каждый. Отсюда Физо сделал правильный вывод, что желтый свет натрия состоит из двух компонент приблизительно равной интенсивности. Максимумы видности полос наблюдались там, где разность хода равнялась целому кратному длины волны каждой компоненты и, следовательно, эти длины волн относились примерно, как 981/980. Физо удалось подтвердить свое заключение прямым наблюдением с призменным спектроскопом.

Позднее более тщательные наблюдения были проведены Майкельсоном [42], определившим с помощью своего интерферометра видность кольцевых полос путем сравнения их с системой кольцевых полос с известной переменной видностью. Это позволило ему построить *кривые видности* полос как функцию разности хода для света целого ряда спектральных линий.

Посмотрим теперь, как кривая видности связана со спектральным распределением интенсивности. Предположим для простоты, что интенсивности двух интерферирующих пучков одинаковы. Для оптической разности хода $\Delta\mathcal{S}$ разность фаз равна

$$\delta(k_0, \Delta\mathcal{S}_1) = k_0 \Delta\mathcal{S}, \tag{108}$$

где $k_0 = 2\pi/\lambda_0$ — волновое число. Из (7.2.17) найдем для интенсивности, обусловленной компонентами с волновыми числами в интервале dk_0,

$$i(k_0, \Delta\mathcal{S})\, dk_0 = 2i_1(k_0)\{1 + \cos(k_0\Delta\mathcal{S})\}\, dk_0, \tag{109}$$

где $i_1(k_0)$ — спектральное распределение интенсивности в каждом пучке. Как уже было показано, отдельные спектральные компоненты складываются некогерентно, и, следовательно, результирующая интенсивность в интерференционной картине равна

$$I(\Delta\mathcal{S}) = 2\int i_1(k_0)\{1 + \cos(k_0\Delta\mathcal{S})\}\, dk_0. \tag{110}$$

Для света спектральной линии $i_1(k_0)$ пренебрежимо мало всюду, за исключением небольшой области k_0 около некоторого среднего волнового числа $\overline{k}_0$. Если теперь принять

$$x = k_0 - \overline{k}_0, \qquad j(x) = i_1(\overline{k}_0 + x), \tag{111}$$

то вместо (110) можно написать

$$I = 2\int j(x)\{1 + \cos[(\overline{k}_0 + x)\Delta\mathcal{S}]\}\, dx. \tag{112}$$

Дальнейший анализ будет вестись так же, как и на стр. 256. Следовательно, последнее соотношение можно переписать (см. (7.3.38)) в виде

$$I(\Delta\mathcal{S}) = P + C(\Delta\mathcal{S})\cos(\overline{k}_0\Delta\mathcal{S}) - S(\Delta\mathcal{S})\sin(\overline{k}_0\Delta\mathcal{S}), \tag{113}$$

где

$$\left.\begin{aligned} P &= 2\int j(x)\, dx, \\ C(\Delta\mathcal{S}) &= 2\int j(x)\cos(x\Delta\mathcal{S})\, dx, \\ S(\Delta\mathcal{S}) &= 2\int j(x)\sin(x\Delta\mathcal{S})\, dx. \end{aligned}\right\} \tag{114}$$

Так как для спектральной линии $j(x)$ не равно нулю только при $|x| \leqslant \bar{k}_0$, то изменения C и S пренебрежимо малы по сравнению с изменениями $\cos(\bar{k}_0 \Delta\mathscr{S})$

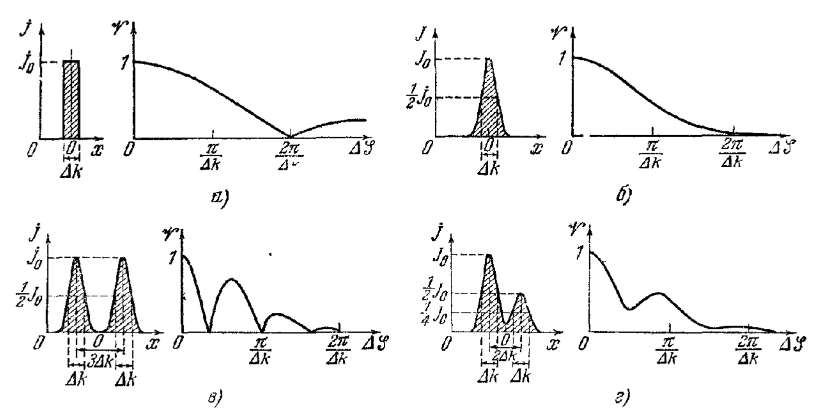


Рис. 7.54. Кривые видности, соответствующие различным спектральным распределениям интенсивности

а) $j = j_0$ когда $|x| \leqslant \frac{1}{2}\Delta k$, $j = 0$, когда $|x| > \frac{1}{2}\Delta k$, $\mathscr{V} = \dfrac{\left|\sin\left(\frac{1}{2}\Delta k \Delta\mathscr{S}\right)\right|}{\left|\frac{1}{2}\Delta k \Delta\mathscr{S}\right|}$; б) $j = j_0 \exp(-\alpha^2 x^2)$, $\mathscr{V} \approx \exp\left[-\left(\frac{\Delta\mathscr{S}}{2\alpha}\right)^2\right]$, в) $j = j_0 \exp[-(\alpha x + \beta)^2] + j_0 \exp[-(\alpha x - \beta)^2]$, $\mathscr{V} \approx \exp\left[-\left(\frac{\Delta\mathscr{S}}{2\alpha}\right)^2\right] \times \left|\cos\left(\frac{\beta}{\alpha}\Delta\mathscr{S}\right)\right|$ при $\frac{\beta}{\alpha\Delta k} = \frac{3}{2}$, г) $j = j_0 \exp[-(\alpha x + \beta)^2] + \frac{1}{2} j_0 \exp[-(\alpha x - \beta)^2]$, $\mathscr{V} \approx \frac{1}{3} \exp\left[-\left(\frac{\Delta\mathscr{S}}{2\alpha}\right)^2\right] \sqrt{5 + 4\cos\left(\frac{2\beta}{\alpha}\Delta\mathscr{S}\right)}$ при $\frac{\beta}{\alpha\Delta k} = 1$. Аналитические выражения для $\mathscr{V}$ выведены в предположении, что свет квазимонохроматичен. На рис. б), в), г) $\Delta k = 2\sqrt{\ln 2}/\alpha \approx 1{,}66/\alpha$.

и $\sin(k_0 \Delta\mathscr{S})$, и отсюда следует, что положение экстремумов I с хорошим приближением определяется соотношением

$$\frac{\partial I}{\partial(\Delta\mathscr{S})} = -\bar{k}_0 [C \sin(\bar{k}_0 \Delta\mathscr{S}) + S \cos(\bar{k}_0 \Delta\mathscr{S})] = 0,$$

т. е.

$$\operatorname{tg}(\bar{k}_0 \Delta\mathscr{S}) = -S/C. \tag{115}$$

Из (113) и (115) для экстремальных величин I найдем

$$I_{\text{экстр}} = P \pm \left|\sqrt{C^2 + S^2}\right|. \tag{116}$$

Отсюда кривая видности описывается соотношением

$$\mathscr{V}(\Delta\mathscr{S}) = \frac{I_{\text{макс}} - I_{\text{мин}}}{I_{\text{макс}} + I_{\text{мин}}} = \frac{\left|\sqrt{C^2 + S^2}\right|}{P}. \tag{117}$$

Отметим, что (113) можно также записать в виде

$$I = P\left[1 + \frac{\left|\sqrt{C^2 + S^2}\right|}{P} \cos(\varphi + \bar{k}_0 \Delta\mathscr{S})\right], \tag{118}$$

где $\operatorname{tg}\varphi = S/C$. Таким образом, кривая видности представляет собой огибающую нормированной кривой интенсивности I/P.

Рассчитанные кривые видности для нескольких возможных спектральных распределений интенсивности показаны на рис. 7.54. Однако на практике мы

встречаемся с обратной задачей, а именно с определением спектрального распределения из наблюдаемой кривой видности. Если функция $j(x)$ симметрична, то $S=0$, и (117) сводится к

$$\mathcal{V} = \frac{|C|}{P}. \qquad (119)$$

В таком случае кривая видности определяет C (не считая постоянного множителя пропорциональности P и знака, который обычно находят из физических соображений) и $j(x)$ получаются из (114) с помощью обратной теоремы Фурье. Однако в общем случае кривая видности позволяет найти только $|\sqrt{C^2+S^2}|$, но этого, вообще говоря, недостаточно для определения $j(x)$, и нужно знать отдельно и C, и S. Рэлей [43] отметил (см. также [44]), что можно определить и C, и S, если измерять не только видность полос, но и их положение, так как последнее с помощью (115) дает отношение C/S; однако такие измерения очень трудны *).

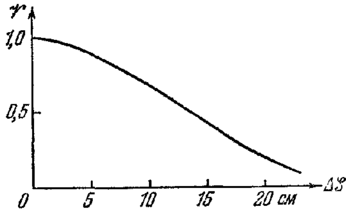


Рис. 7.55. Кривая видности для красной линии кадмия [46].

Несмотря на эти трудности, Майкельсон смог выяснить структуру простых спектральных линий, получившую убедительное подтверждение в последующих исследованиях. В частности, он нашел, что красная линия кадмия ($\lambda = 6438$ Å) наиболее близка к монохроматичной из всех исследованных им линий; ее кривая видности (рис. 7.55) соответствует симметричному спектральному распределению гауссовой формы с полушириной, равной лишь 0,013 Å; ему удалось также наблюдать интерференционные полосы с разностью хода, превышающей 500 000 длин волн (~30 *см*).

Такой метод анализа спектров с помощью двухлучевой интерферометрии имеет исторический интерес как первое применение интерференции в спектроскопии; позднее он был вытеснен методами многолучевой интерферометрии (см. § 7.6). Однако сравнительно недавно снова возродился интерес к двухлучевому методу в применении к инфракрасной области спектра (см., например, [47] и обзорную статью по Фурье спектроскопии [47a]), так как в этом случае он имеет ряд технических преимуществ.

§ 7.6. Многолучевая интерференция

При падении пучка света на прозрачную пластинку на ее поверхностях происходят многократные отражения, в результате чего с каждой стороны пластинки выходит ряд пучков с убывающей амплитудой. Рассматривая интерференционные эффекты, возникающие с такими пластинками (см. пп. 7.5.1, 7.5.2, 7.5.6), мы пренебрегали вкладом в результирующую интенсивность пучков, испытавших больше двух отражений. Такое допущение оправдано при малой отражательной способности поверхностей. Теперь мы учтем все отраженные пучки и покажем, что при большой отражательной способности поверхностей распределение интенсивности в картине полос изменяется. Именно поэтому многолучевая интерференция находит важное практическое применение.

7.6.1. Многолучевые интерференционные полосы, полученные с плоскопараллельной пластинкой. Рассмотрим плоскопараллельную прозрачную пластинку с показателем преломления n', находящуюся в среде с показателем преломления n, и предположим, что на эту пластинку под углом θ падает плоская волна монохроматического света. Пусть луч SB_1 (рис. 7.56) представляет

*) Большое число таких измерений описано в работе Перарда [45].

направление распространения падающей волны. На первой поверхности эта волна разделяется на две плоские волны: одну, отраженную в направлении B_1C_1, и другую, прошедшую в пластинку в направлении B_1D_1. Прошедшая волна падает на вторую поверхность под углом θ' и здесь снова разделяется на две плоские волны: одну, прошедшую в направлении D_1E_1, и другую, отраженную обратно в пластинку в направлении D_1B_2. Такой процесс деления волны, остающейся в пластинке, продолжается, как показано на рисунке.

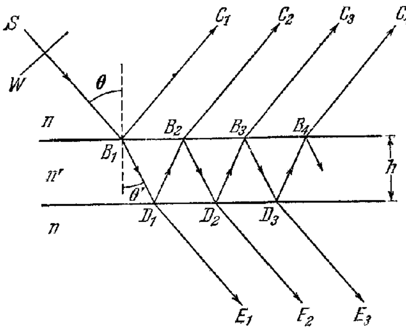


Рис 7 56 Отражение плоской волны в плоскопараллельной пластинке.

Пусть $A^{(i)}$ — амплитуда электрического вектора падающей волны, которую мы считаем линейно поляризованной либо параллельно, либо перпендикулярно плоскости падения. Как и в п. 1.5 2, мы полагаем, что $A^{(i)}$ комплексна и ее фаза равна постоянной части фазы соответствующей волновой функции. Для каждого члена совокупности отраженных или прошедших волн переменная часть фазы волновой функции отличается от такой же части фазы предыдущего члена на величину, соответствующую двукратному прохождению луча в пластинке. Согласно (7.5.6) эта разность фаз равна

$$\delta = \frac{4\pi}{\lambda_0} n'h\cos\theta', \tag{1}$$

где h — толщина пластинки, а λ_0— длина волны в вакууме. Пусть для волны, идущей из окружающей среды в пластинку, r — коэффициент отражения (отношение амплитуд отраженной и падающей волн), а t — коэффициент пропускания (отношение амплитуд прошедшей и падающей волн), пусть далее r' и t'— соответствующие коэффициенты для волны, идущей из пластинки в окружающую среду. Комплексные амплитуды волн, отраженных от пластинки, запишутся тогда в виде

$$rA^{(i)},\ tt'r'\exp(i\delta),\ tt'r'^3A^{(i)}\exp(2i\delta),\ \ldots,\ tt'r'^{(2p-3)}A^{(i)}\exp(i(p-1)\delta).$$

Аналогично комплексные амплитуды волн, прошедших сквозь пластинку (отбрасывая несущественный постоянный фазовый множитель), будут равны

$$tt'A^{(i)},\ tt'r'^2A^{(i)}\exp(i\delta),\ tt'r'^4A^{(i)}\exp(2i\delta)\ldots tt'r'^{2(p-1)}A^{(i)}\exp(i(p-1)\delta),\ \ldots$$

Величины r, r', t и t' связаны с n, n', θ и θ' формулами Френеля (см п. 5.2). Сейчас нам не нужны они в явном виде, а важно только существование таких связей. Таким образом, из соотношений (1.5.20а) и (1.5.35) видно, что для каждой поляризованной компоненты

$$tt' = \mathcal{T}; \tag{2}$$

аналогично из уравнения (1.5.21а) получим

$$r = -r', \tag{3}$$

откуда, согласно (1.5.33),

$$r^2 = r'^2 = \mathcal{R}, \tag{4}$$

где $\mathcal{R}$ и $\mathcal{T}$ — соответственно отражательная и пропускательная способности поверхностей пластинки — связаны соотношением

$$\mathcal{R} + \mathcal{T} = 1. \tag{5}$$

При суперпозиции первых p отраженных волн амплитуда $A^{(r)}(p)$ электрического вектора отраженного света дается выражением

$$A^{(r)}(p)=$$
$$=\{r+tt'r'\exp(i\delta)(1+r'^2\exp(i\delta)+\ldots+r'^{2(p-2)}\exp[i(p-2)\delta]\}A^{(i)}=$$
$$=\left\{r+\left(\frac{1-r'^{2(p-1)}\exp[i(p-1)\delta]}{1-r'^2\exp(i\delta)}\right)tt'r'\exp(i\delta)\right\}A^{(i)}. \quad (6)$$

Если пластинка достаточно длинна, число отраженных волн велико, и в пределе при $p\to\infty$ мы получим из (6) и (3)

$$A^{(r)}\equiv A^{(r)}(\infty)=-\frac{r'\{1-(r'^2+tt')\exp(i\delta)\}}{1-r'^2\exp(i\delta)}A^{(i)}. \quad (7)$$

Из (2), (4) и (5) находим

$$A^{(r)}=\frac{(1-e^{i\delta})\sqrt{\mathcal{R}}}{1-e^{i\delta}}A^{(i)}, \quad (8)$$

и, следовательно, интенсивность $I^{(r)}=A^{(r)}A^{(r)*}$ отраженного света равна

$$I^{(r)}=\frac{(2-2\cos\delta)\mathcal{R}}{1+\mathcal{R}^2-2\mathcal{R}\cos\delta}I^{(i)}=\frac{4\mathcal{R}\sin^2\frac{\delta}{2}}{(1-\mathcal{R})^2+4\mathcal{R}\sin^2\frac{\delta}{2}}I^{(i)}, \quad (9)$$

где $I^{(i)}=A^{(i)}A^{(i)*}$ — интенсивность падающего света.

Таким же путем получается выражение для амплитуды $A^{(t)}$ прошедшего света, а именно

$$A^{(t)}(p)=tt'(1+r'^2\exp(i\delta)+\ldots$$
$$+r'^{2(p-1)}\exp[i(p-1)\delta])A^{(i)}=$$
$$=\left(\frac{1-r'^{2p}\exp(ip\delta)}{1-r'^2\exp(i\delta)}\right)tt'A^{(i)}. \quad (10)$$

В пределе при $p\to\infty$ (10) сводится к

$$A^{(t)}\equiv A^{(t)}(\infty)=\frac{tt'}{1-r'^2e^{i\delta}}A^{(i)}. \quad (11)$$

Отсюда, используя (2) и (4), получим

$$A^{(t)}=\frac{\mathcal{T}}{1-\mathcal{R}e^{i\delta}}A^{(i)}, \quad (12)$$

и соответствующая интенсивность $I^{(t)}=A^{(t)}A^{(t)*}$ прошедшего света равна

$$I^{(t)}=\frac{\mathcal{T}^2}{1+\mathcal{R}^2-2\mathcal{R}\cos\delta}I^{(i)}=$$
$$=\frac{\mathcal{T}^2}{(1-\mathcal{R})^2+4\mathcal{R}\sin^2\frac{\delta}{2}}I^{(i)}. \quad (13)$$

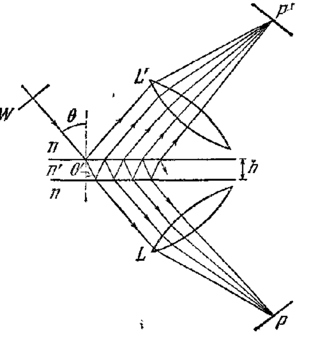


Рис. 7.57. К возникновению многолучевых интерференционных полос равного наклона в случае плоскопараллельной пластинки.

Формулы (9) и (13), известные под названием *формул Эйри*, находятся в согласии с результатами, полученными ранее из общей теории распространения волн в слоистой среде (см. § 1.6); если в уравнениях (1.6.60) положить $r_{12}=r$, $r_{23}=r'$, $t_{12}=t$, $t_{23}=t'$, $2\beta=\delta$ и использовать соотношения (2) и (3), то получим (9) и (13) *).

Предположим теперь, что плоские волны равной интенсивности падают на пластинку под разными, мало различающимися углами и прошедший свет

*) При этом сопоставлении следует помнить, что в отличие от обозначений, принятых в настоящем разделе, символы $\mathcal{R}$ и $\mathcal{T}$ в § 1.6 обозначают отражательную и пропускательную способности всей пластинки.

собирается линзой L (рис. 7.57). В точке P фокальной плоскости L интенсивность прошедшего света относится к интенсивности в P в отсутствие пластинки, как $I^{(t)}/I^{(i)}$; следовательно, согласно (13), в присутствии пластинки максимумы интенсивности в P будут тогда, когда порядок интерференции m, определяемый выражением

$$m=\frac{\delta}{2\pi}=\frac{2n'h\cos\theta'}{\lambda_0}, \tag{14}$$

равен целым числам 1, 2, ..., минимумы,— когда он равен полуцелым числам $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ Очевидно, что в фокальной плоскости L появляются полосы равного наклона, соответствующие местам с постоянным θ' (и, следовательно, θ). Аналогично, если свет, отраженный от пластинки, собирается линзой L', то в ее фокальной плоскости также появятся линии равного наклона. Из (9) следует, что максимумы интенсивности в этой картине для отраженного света соответствуют полуцелым значениям порядка интерференции $m=\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$, а минимумы — целым числам 1, 2... Следовательно, положение полос в обеих картинах совпадает с тем, которое получалось при приближенном рассмотрении в п. 7.5.1, где с каждой стороны пластинки мы учитывали только два первых пучка.

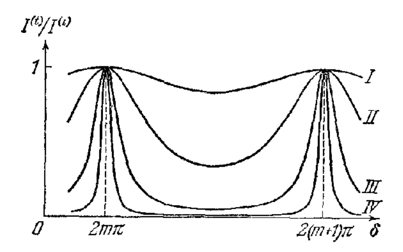


Рис. 7.58. Многолучевые интерференционные полосы равного наклона в прошедшем свете.

Отношение $I^{(t)}/I^{(i)}$ интенсивности прошедшего света к интенсивности падающего как функция разности фаз δ, m — целое число. *I*) $F=0,2$, $\mathscr{R}=0,046$; *II*) $F=2$, $\mathscr{R}=0,27$; *III*) $F=20$, $\mathscr{R}=0,64$; *IV*) $F=200$, $\mathscr{R}=0,87$.

Распределение интенсивности в интерференционных картинах в отраженном и прошедшем свете описывается (9) и (13) и с помощью (5) их можно представить в виде

$$\frac{I^{(r)}}{I^{(i)}}=\frac{F\sin^2(\delta/2)}{1+F\sin^2(\delta/2)}, \tag{15а}$$

$$\frac{I^{(t)}}{I^{(i)}}=\frac{1}{1+F\sin^2(\delta/2)}, \tag{15б}$$

где параметр F определяется соотношением

$$F=\frac{4\mathscr{R}}{(1-\mathscr{R})^2}. \tag{16}$$

Очевидно, обе картины дополнительны в том смысле, что

$$\frac{I^{(r)}}{I^{(i)}}+\frac{I^{(t)}}{I^{(i)}}=1. \tag{17}$$

График зависимости $I^{(t)}/I^{(i)}$ от разности фаз δ при разных значениях F показан на рис. 7.58. Когда $\mathscr{R}$ мало по сравнению с единицей, F также мало по сравнению с единицей, и мы можем разложить $1\Big/\left(1+F\sin^2\frac{\delta}{2}\right)$ в ряд (см. (15)) и сохранить члены с F только в первой степени. Это дает

$$\frac{I^{(r)}}{I^{(i)}}\approx F\sin^2\frac{\delta}{2}=\frac{F}{2}(1-\cos\delta), \tag{18а}$$

$$\frac{I^{(t)}}{I^{(i)}}\approx 1-F\sin^2\frac{\delta}{2}=1-\frac{F}{2}(1-\cos\delta), \tag{18б}$$

т. е. соотношения для изменения интенсивности имеют вид (7.2.15), типичный для двух интерферирующих пучков. Если $\mathscr{R}$ увеличивается, приближаясь к единице так, что и F становится большим, то интенсивность минимумов картины в прошедшем свете падает и максимумы становятся резче. Интенсивность в прошедшем свете всюду, за исключением участков, находящихся в непосред-

ственной близости к максимумам, очень мала. Таким образом, интерференционная картина в прошедшем свете имеет вид узких светлых полос на почти совершенно темном фоне. В отраженном свете она имеет вид узких темных полос на почти равномерном светлом фоне. Резкость полос принято измерять *полушириной интенсивности*, или просто *полушириной*. Для картины в прошедшем свете она равна расстоянию между точками, лежащими по обе стороны максимума в том месте, где интенсивность уменьшается до половины максимальной величины. Отношение расстояния между соседними полосами к полуширине мы назовем *резкостью* полос и обозначим буквой $\mathcal{F}$. У полосы целого порядка m точки, где интенсивность равна половине максимальной величины, находятся при

$$\delta = 2m\pi \pm \varepsilon/2, \tag{19}$$

где в соответствии с (15б)

$$\frac{1}{1+F\sin^2\frac{\varepsilon}{4}} = \frac{1}{2}. \tag{20}$$

Когда F достаточно велико, ε настолько мало, что в (20) мы можем принять $\sin(\varepsilon/4) = \varepsilon/4$, и получим для полуширины

$$\varepsilon = \frac{4}{\sqrt{F}}. \tag{21}$$

Так как расстояние между соседними полосами соответствует изменению δ на 2π, то для резкости находим

$$\mathcal{F} = \frac{2\pi}{\varepsilon} = \frac{\pi\sqrt{F}}{2}. \tag{22}$$

До сих пор мы предполагали, что свет строго монохроматичен. В случае квазимонохроматического света распределение интенсивности равно сумме распределений интенсивностей типа (15), обусловленных каждой монохроматической компонентой; если эти компоненты занимают область длин волн $\Delta\lambda_0$ вблизи средней длины волны $\bar{\lambda}_0$, то максимумы порядка m распределены в области, соответствующей $|\Delta\delta|$ в картине, получающейся в длине волны $\bar{\lambda}_0$. Согласно (14) и пренебрегая зависимостью $n'h$ от длины волны, находим $|\Delta\delta| = 2\pi m\Delta\lambda_0/\bar{\lambda}_0$; можно считать, что интерференционные картины, создаваемые этими компонентами, совпадают, а результирующее распределение интенсивности оказывается таким же, как и в случае строго монохроматического света с длиной волны $\bar{\lambda}_0$, если $|\Delta\delta|$ пренебрежимо мало по сравнению с полушириной ε монохроматической полосы. В соответствии с (22) последнее условие выполняется при $m\mathcal{F} \ll \bar{\lambda}_0/\Delta\lambda_0$, т. е. при

$$\mathcal{F}\Delta\mathcal{S} \ll \frac{\bar{\lambda}_0^2}{\Delta\lambda_0}, \tag{23}$$

где

$$\Delta\mathcal{S} = \frac{\bar{\lambda}_0}{2\pi}\delta = m\bar{\lambda}_0 \tag{24}$$

— оптическая разность хода между соседними интерферирующими пучками. Неравенство (23) аналогично неравенству (7.3.15), относящемуся к случаю интерференции двух пучков. Величина, стоящая в правой части неравенства (23), тождественна длине когерентности света (см. (7.5.107)).

Из предыдущего ясно, что при возрастании отражательной способности поверхностей и, следовательно, при увеличении $\mathcal{F}$ распределение интенсивности становится более благоприятным для определения положения полос. В интерференционной картине в проходящем свете полосы, принадлежащие различным монохроматическим компонентам, разделяются более четко. По этим

причинам многолучевая интерференция имеет бо́льшое практическое значение. В оптическом диапазоне отражательная способность поверхности между двумя диэлектриками при нормальном падении мала; например, для границы воздух ($n\sim 1$) — стекло ($n'\sim 1{,}5$) находим из (1.5.37) $\mathcal{R}\sim 0{,}04$. Отражение увеличивается при косом падении и, как мы видели в § 1.5, $\mathcal{R}\to 1$, когда в оптически более плотной среде свет падает под углом, близким к критическому. Отражательную способность, мало отличающуюся от единицы, можно получить и при почти нормальном падении света, путем нанесения на поверхность диэлектрика многослойного покрытия из других подходящих диэлектриков (см. § 1.6) или путем нанесения частично прозрачного слоя металла. Теория отражения от металлических пленок на диэлектрической подложке изложена в гл. 13, но здесь, забегая вперед, можно сказать, что такие слои поглощают свет, что сдвиг фаз при отражении не должен быть обязательно равен нулю или π и что отражательные способности и сдвиги фаз на обеих поверхностях слоя различны, если у граничащих с ними диэлектриков неодинаковые показатели преломления. Следовательно, приведенный выше анализ неприменим в случае пластинок, покрытых металлическими слоями. Однако при идентичных слоях на обеих поверхностях пластинки уравнения (12) и (13) остаются верными, если под $\mathcal{R}$ мы будем подразумевать отражательную способность при внутреннем отражении и заменим δ, определенную соотношением (1), на величину

$$\delta=\frac{4\pi}{\lambda_0}n'h\cos\theta'+2\varphi, \tag{25}$$

где φ — сдвиг фазы при внутреннем отражении. Полагая

$$\mathcal{R}+\mathcal{T}+\mathcal{A}=1, \tag{26}$$

где $\mathcal{A}$ — часть света, поглощенная металлом, получим из (13) и (16) соотношение

$$\frac{I^{(t)}}{I^{(i)}}=\left(1-\frac{\mathcal{A}}{1-\mathcal{R}}\right)^2\frac{1}{1+F\sin^2(\delta/2)}. \tag{27}$$

Сравнивая (27) и (15б), мы видим, что для данного F поглощение уменьшает интенсивность картины в проходящем свете на множитель *) $[1-\mathcal{A}/(1-\mathcal{R})]^{-2}$. При падении света в направлении, близком к нормальному, изменение фазы φ при отражении эквивалентно увеличению оптической толщины пластинки на $\varphi\lambda_0/2\pi$; при косом падении возникает и другое осложнение, а именно то, что сдвиг фаз зависит от поляризации света.

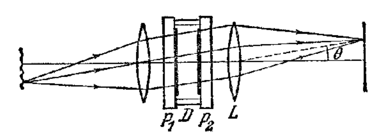


Рис. 7.59. Интерферометр Фабри — Перо.

7.6.2. Интерферометр Фабри — Перо. Многолучевые интерференционные полосы, создаваемые плоскопараллельной пластинкой при почти нормальном освещении, используются в *интерферометре Фабри Перо* [49]. Основными частями его служат две стеклянные или кварцевые пластины P_1 и P_2 (рис. 7.59) с плоскими поверхностями. Внутренние поверхности пластин, покрытые частично прозрачными пленками с высокой отражательной способностью, параллельны, и воздух, заключенный между этими поверхностями, образует плоскопараллельную пластинку. Сами пластины делают слегка клиновидными, чтобы устранить вредное влияние света, отраженного внешними непокрытыми поверхностями. В первых образцах прибора одна пластина была неподвижна, а другая устанавливалась на салазках, что позволяло перемещать ее с помощью винта относительно первой. Однако вследствие ненадежности механической конструкции такие системы вышли из употребления. В настоящее время пла-

*) При многослойных диэлектрических отражающих покрытиях распределение интенсивности в интерференционной картине в прошедшем свете тоже определяется (27), а уменьшение интенсивности в этом случае обусловлено в основном рассеянием света (см. [48]).

стины разделяют неподвижным кольцом D из инвара или кварца с тремя выступами на торцах, к которым пластины прижимаются пружинами. Кольца обработаны с большой точностью, так что положение плоскостей, заданное выступами, максимально близко к параллельному, а тонкая регулировка осуществляется изменением нажима пружин. Интерферометр такого типа с фиксированным расстоянием между пластинами иногда называют *эталоном Фабри — Перо*.

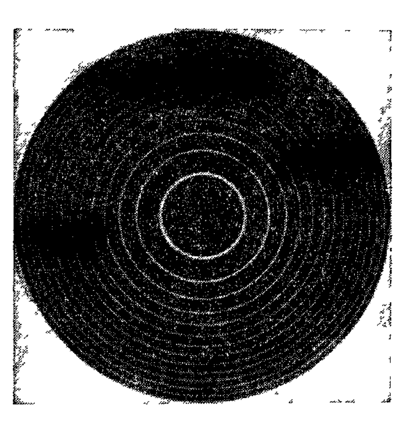


Рис. 7.60. Полосы, полученные с интерферометром Фабри — Перо.

Как мы видели в п. 7.6.1, свет от протяженного квазимонохроматического источника S, удовлетворяющий условию (23), образует узкие световые полосы равного наклона в фокальной плоскости линзы L. Согласно (25) порядок интерференции равен

$$m=\frac{\delta}{2\pi}=\frac{2n'h\cos\theta'}{\lambda_0}+\frac{\varphi}{\pi}, \qquad (28)$$

где n' — показатель преломления воздуха между пластинами, h — расстояние между отражающими поверхностями, θ' — угол отражения, а φ — сдвиг фазы. Ось линзы обычно нормальна к пластинам, поэтому светлые полосы, соответствующие целым значениям m для света, прошедшего в направлении нормали к пластинам, имеют вид окружностей с общим центром в фокусе линзы (рис. 7.60). В этой точке m имеет максимальную величину m_0, равную

$$m_0=\frac{2n'h}{\lambda_0}+\frac{\varphi}{\pi}. \qquad (29)$$

В общем случае m_0 отлично от целого числа, и мы можем написать

$$m_0=m_1+e, \qquad (30)$$

где m_1 — целый порядок внутреннего наиболее светлого кольца, а e — величина, меньшая единицы,— дробный порядок в центре. Из (28) — (30) тем же путем, каким выводилось уравнение (7.5.13), можно получить угловой радиус θ_p p-го светлого кольца от центра (если θ_p не слишком велико) в виде

$$\theta_p=\frac{1}{n}\sqrt{\frac{n'\lambda_0}{h}}\sqrt{p-1+e}, \qquad (31)$$

где n — показатель преломления воздуха вне пластин. Диаметр D_p этого кольца равен, следовательно,

$$D_p^2=(2f\theta_p)^2=\frac{4n'\lambda_0 f^2}{n^2h}(p-1+e); \qquad (32)$$

здесь f — фокусное расстояние линзы L.

Важнейшими практическими характеристиками интерферометра Фабри — Перо служат резкость $\mathcal{F}$, определенная выше как отношение расстояния между полосами к их полуширине, *максимальное пропускание*

$$\tau=\left(\frac{I^{(t)}}{I^{(i)}}\right)_{\text{макс}} \qquad (33)$$

и *контрастность*

$$\mathcal{C}=\left(\frac{I^{(t)}}{I^{(i)}}\right)_{\text{макс}}\bigg/\left(\frac{I^{(t)}}{I^{(i)}}\right)_{\text{мин}}, \qquad (34)$$

где $I^{(t)}$ — интенсивность в некоторой точке интерференционной картины, а $I^{(i)}$ — соответствующая интенсивность при отсутствии интерферометра. Если мы допустим теперь, что внутренние поверхности пластин плоски и параллельны и пренебрежем отражениями на внешних поверхностях, то $I^{(t)}/I^{(i)}$ определится (27). Сравнивая (27) и (15б), мы видим, что резкость задается выражением (22), а используя (16), найдем

$$\mathcal{F} = \frac{\pi\sqrt{\mathcal{R}}}{1-\mathcal{R}}. \tag{35}$$

Согласно (27) максимум пропускания равен

$$\tau = \left(1 - \frac{\mathcal{A}}{1-\mathcal{R}}\right)^2, \tag{36}$$

согласно (16), (27) и (35) контрастность записывается в виде

$$\mathcal{C} = 1 + F = \left(\frac{1+\mathcal{R}}{1-\mathcal{R}}\right)^2 = 1 + \frac{4\mathcal{F}^2}{\pi^2}. \tag{37}$$

Как мы отмечали в п. 7.6.1, отражающими покрытиями пластин могут служить либо металлические пленки (чаще всего из серебра и алюминия),

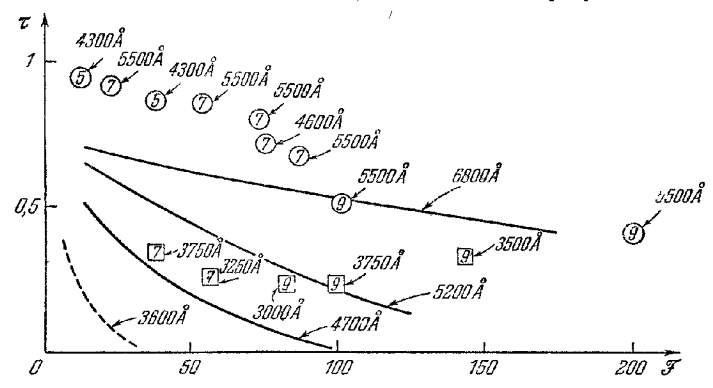


Рис. 7.61. Максимальное пропускание τ и резкость $\mathcal{F}$ при различных отражающих покрытиях.

Предполагается, что пластины идеально плоски и их поверхности строго параллельны. Сплошные кривые — *свежеприготовленные серебряные пленки* [56], пунктирная кривая — *алюминиевые пленки* [57], кружочки — *пленки из сульфида цинка — криолита* [58], квадратики — *пленки из хлористого свинца — фтористого магния* [59].

Для диэлектрических пленок указаны число слоев и оптимальная длина волны

либо диэлектрические пленки из чередующихся слоев материалов с высоким и малым показателем преломления (например, сульфида цинка и криолита) и с оптической толщиной $\lambda_0/4$ каждый *). Пленки обоих типов наносятся термическим испарением в вакууме. Для данной длины волны $\mathcal{R}$, вообще говоря, возрастает с увеличением толщины пленки из металла и с увеличением числа слоев диэлектрика (см. табл. 1.3). Однако было установлено, что для покрытий обоих типов при больших величинах $\mathcal{R}$, представляющих практический интерес, увеличение $\mathcal{R}$ сопровождается увеличением величины $\mathcal{A}/(1-\mathcal{R})$, т. е. согласно (35) и (36), τ уменьшается с увеличением $\mathcal{F}$. Таким образом, получить одновременно большие величины максимального пропускания и резкости (или контрастности) невозможно, и практически приходится находить какой-то компромисс в выборе этих величин. На рис. 7.61 приведены величины τ и $\mathcal{F}$, полученные из измерения на типичных пленках. Как мы видим (исключая крайнюю красную часть спектра), наибольшие величины резкости при данном максимальном пропускании достигаются с диэлектрическими покрытиями.

*) Применяется также комбинация металлических и диэлектрических пленок [50].

Следует отметить, однако, что коэффициент отражения многослойных диэлектрических покрытий высок только в ограниченной области длин волн вблизи длины волны λ_0, для которой оптическая толщина каждого слоя составляет $\lambda_0/4$, и поэтому диэлектрические покрытия не следует применять там, где один интерферометр используется для широкого спектрального интервала. В видимом участке спектра серебряные покрытия дают более резкие полосы для данного максимального пропускания, чем алюминий, тогда как в области меньших длин волн (ниже 4000 Å) алюминиевые покрытия лучше и могут применяться в ультрафиолете вплоть до 2000 Å.

Предыдущее обсуждение относилось к идеальному интерферометру с идеально плоскими и параллельными отражающими поверхностями *) Практически же поверхности интерферометрических пластин невозможно сделать абсолютно плоскими и, следовательно, расстояние h между ними всегда изменяется по площади пластин. Влияние такого непостоянства h рассматривалось Дюфуром и Пикка [54] и Шаббалом [55]. Они показали, что резкость и максимальное пропускание всегда меньше, чем это следует из (35) и (36), и что при $\mathcal{R} \to 1$ резкость приближается к пределу $\mathcal{F}_d$, зависящему только от дефектов пластин, т. е. для данных пластин существует верхний предел резкости полос, который нельзя превысить ни при какой отражательной способности покрытий. Величина $\mathcal{F}_d$ зависит от формы и величины отступления от плоскопараллельности. В частности, когда дефект пластины заключается в небольшой ее сферичности, такой, что $n'h$ изменяется на λ_0/q от центра отверстия интерферометра к его краю, то $\mathcal{F}_d = q/2$. В качестве примера рассмотрим интерферометр с $q = 100$ (т. е. $\mathcal{F}_d = 50$), пластины которого покрыты свежеприготовленными пленками серебра; пусть длина волны используемого света примерно равна 5200 Å. Сравнение резкости $\mathcal{F}$ и максимального пропускания τ, получаемых с таким интерферометром, с их величинами, указанными на рис. 7.61, проведено в табл. 7.1. Мы видим, что при $\mathcal{F} \gtrsim \mathcal{F}_d$ увеличение отражательной способности приводит к уменьшению максимального пропускания, слегка компенсируемому увеличением резкости. Этот пример показывает также, с какой высокой степенью точности должна изготавливаться плоскость пластин, чтобы можно было использовать все преимущества высокой отражательной способности.

Таблица 7.1

Резкость $\mathcal{F}$ и максимальное пропускание τ для плоских пластин (а) и для пластин со сферической кривизной (б), у которых различие величин $n'h$ в центре и на краю составляет 52 Å

а) Плоские пластины		б) Пластины со сферической кривизной ($\mathcal{F}_d = 50$)	
$\mathcal{F}$	τ	$\mathcal{F}$	τ
25	0,59	22	0,55
50	0,44	36	0,34
75	0,30	42	0,20
100	0,20	45	0,11
125	0,13	46	0,06

Величины для (а) взяты из рис. 7.61. Величины для (б) получены из теоретических кривых [55].

7.6.3. Применение интерферометра Фабри — Перо для изучения тонкой структуры спектральных линий. При освещении интерферометра Фабри — Перо квазимонохроматическим светом, не удовлетворяющим условиям (23), распределение интенсивности в прошедшем свете отличается от даваемого

*) Другой тип интерферометра Фабри — Перо, в котором плоские зеркала заменены сферическими зеркалами с равными радиусами кривизны и установлены так, что их фокусы совпадают, описан в [51]. Интерферометры такого типа применяются, например, в качестве резонаторов в лазерах [52, 53].

выражением (27) и содержит некоторую информацию о спектральном распределении используемого света. В частности, допустим, что свет состоит из двух монохроматических компонент. Представим себе, что различие в их длинах волн постепенно возрастает, и если они не отличаются слишком сильно по интенсивности, наличие двух компонент в конце концов будет замечено, так как в интерференционной картине появятся две смещенные друг относительно друга системы максимумов. В таком случае говорят, что компоненты *разрешены* интерферометром. Подобным способом Фабри и Перо [49] удалось непосредственно наблюдать тонкую структуру спектральных линий, существование которой Майкельсон только предполагал (см. п. 7.5.8). С тех пор в этой области спектроскопии интерферометр Фабри — Перо играет доминирующую роль.

Для сравнения способностей различных приборов разрешать структуру спектров, удобно рассматривать линию, состоящую из двух компонент равной интенсивности, и установить, быть может несколько произвольно, величину смещения максимумов, при котором компоненты «начнут разрешаться». Пусть $\lambda_0 \pm \frac{1}{2}\Delta\lambda_0$ — длины волн двух таких компонент; тогда величина $\lambda_0/\Delta\lambda_0$ называется *разрешающей силой* прибора. Такой критерий разрешения впервые введен лордом Рэлеем [60] для спектроскопов с призмой или решеткой, у которых распределение интенсивности в монохроматическом свете имеет вид $I(\delta) = \left\{\sin\left(\frac{\delta}{2}\right) / \left(\frac{\delta}{2}\right)\right\}^2 I_{\text{макс}}$. Рэлей предложил считать, что в этом случае две компоненты равной интенсивности начинают разрешаться, когда главный максимум интенсивности одной совпадает с первым минимумом другой (рис. 7.62). При таком комбинированном распределении отношение интенсивности в средней точке к интенсивности в максимуме равно $8/\pi^2 \simeq 0{,}811$.

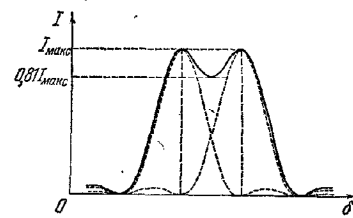


Рис. 7.62. Две спектральные монохроматические компоненты, которые начали разрешаться согласно критерию Рэлея.

Будем считать подобное отношение интенсивности в седловине и в максимуме критерием разрешения в интерферометре Фабри — Перо *). Если отражающая способность пластин не столь велика, чтобы начало сказываться их несовершенство, распределение интенсивности $I(\delta)$, обусловленной одной монохроматической компонентой, можно записать на основании (27) в виде

$$I(\delta) = \frac{I_0}{1 + F\sin^2\frac{\delta}{2}}, \tag{38}$$

а полную интенсивность $I_{\text{полн}}(\delta, \varepsilon)$, создающуюся в результате суперпозиции двух таких компонент, расстояние между которыми соответствует изменению δ на величину ε,— в виде

$$I_{\text{полн}}(\delta, \varepsilon) = I\left(\delta + \frac{1}{2}\varepsilon\right) + I\left(\delta - \frac{1}{2}\varepsilon\right) =$$
$$= \frac{I_0}{1 + F\sin^2\frac{\left(\delta + \frac{1}{2}\varepsilon\right)}{2}} + \frac{I}{1 + F\sin^2\frac{\left(\delta - \frac{1}{2}\varepsilon\right)}{2}}. \tag{39}$$

*) Критерию Рэлея нельзя приписать какой-либо специальный физический смысл, и время от времени предлагаются другие критерии разрешения. При сравнении различных инструментов специальный выбор критерия мало существен.

Полная интенсивность в средней точке между максимумами интенсивности двух компонент равна $I_{\text{полн}}(2m\pi, \varepsilon)$, где m — целое число; если (при условии, что компоненты начинают разрешаться) мы считаем, что максимумы полной интенсивности совпадают с максимумами интенсивности компонент, то полная интенсивность в максимумах равна $I_{\text{полн}}\left(2m\pi \pm \frac{1}{2}\varepsilon, \varepsilon\right)$. Следовательно, используя выбранный нами критерий, мы должны считать, что две линии начинают разрешаться, если ε удовлетворяет соотношению

$$\frac{2I_0}{1+F\sin^2\frac{\varepsilon}{4}} = 0{,}81\left\{I_0+\frac{I_0}{1+F\sin^2\frac{\varepsilon}{2}}\right\}. \tag{40}$$

При большой резкости полос ε в (40) мало по сравнению с $\pi/2$, и можно принять $\sin\varepsilon = \varepsilon$; тогда (40) сведется к

$$F^2\varepsilon^4 - 15{,}5\,F\varepsilon^2 - 30 = 0,$$

откуда, учитывая (22), находим

$$\varepsilon = \frac{4{,}15}{\sqrt{F}} = \frac{2{,}07\pi}{\mathcal{F}}. \tag{41}$$

Далее из (28), принимая n' не зависящим от длины волны и полагая h столь большим, что можно пренебречь φ по сравнению с δ, получим

$$|\Delta\delta| = \frac{2\pi n'h\cos\theta'}{\lambda_0^2}\Delta\lambda_0 = 2\pi m\frac{\Delta\lambda_0}{\lambda_0}. \tag{42}$$

На пределе разрешения $|\Delta\delta|$ равно величине ε, определяемой (41), и, следовательно, для разрешающей силы интерферометра имеем

$$\frac{\lambda_0}{\Delta\lambda_0} = 0{,}97m\mathcal{F}. \tag{43}$$

По аналогии с выражением (8.6.14) для разрешающей силы дифракционной решетки с *конечным* числом интерферирующих пучков *равной* интенсивности множитель $0{,}97\,\mathcal{F}$ иногда называют *эффективным числом пучков*. При падении света, близком к нормальному, $m \approx 2n'h/\lambda_0$, и можно считать, что разрешающая сила равна

$$\frac{\lambda_0}{\Delta\lambda_0} \approx \frac{2\mathcal{F}n'h}{\lambda_0}. \tag{44}$$

Итак разрешающая сила интерферометра пропорциональна резкости и оптическому расстоянию между пластинами. В качестве примера возьмем $\mathcal{F} = 30$ ($\mathcal{R} \sim 0{,}9$), величину легко достижимую при работе в видимом спектре, и $n'h = 4$ *мм*; тогда разрешающая сила для $\lambda_0 - 5000$ А примерно равна $5\cdot 10^5$, т. е. имеет тот же порядок, что и разрешающая сила наибольших штриховых дифракционных решеток. Для спектроскопических целей удобно также ввести интервал $\Delta\kappa_0$ спектроскопических волновых чисел ($\kappa_0 = 1/\lambda_0$), соответствующий наименьшей разрешаемой разности длин волн $\Delta\lambda_0$ и определяемый как

$$\Delta\kappa_0 = \frac{\Delta\lambda_0}{\lambda_0^2} \approx \frac{1}{2\mathcal{F}n'h}. \tag{45}$$

Величину $\Delta\kappa_0$ иногда называют *пределом разрешения* интерферометра. В предыдущем примере предел разрешения составлял примерно $0{,}04$ *см*$^{-1}$.

Если у двух компонент различие в длинах волн достаточно велико, смещение интерференционных картин друг относительно друга больше расстояния между соседними максимумами каждой компоненты. Тогда мы говорим о «перекрытии» порядков. Разность длин волн $(\Delta\lambda_0)_{\text{о.д}}$, соответствующая смещению на один порядок ($|\Delta\delta| = 2\pi$), называется *областью дисперсии* интерферометра.

При падении света, близком к нормальному, находим из (42)

$$(\Delta\lambda_0)_{\text{о. д}}=\frac{\lambda_0}{m}\approx\frac{\lambda_0^2}{2n'h}. \tag{46}$$

Если область дисперсии выразить через спектроскопическое волновое число, то

$$(\Delta\kappa_0)_{\text{о. д}}=\frac{(\Delta\lambda_0)_{\text{о. д}}}{\lambda_0^2}\approx\frac{1}{2n'h}. \tag{47}$$

Как мы видим, размер области дисперсии обратно пропорционален расстоянию между пластинами, и значит, увеличение разрешающей силы, получающееся при увеличении этого расстояния, сопровождается пропорциональным уменьшением области дисперсии. Сравнивая (44) и (46), мы также видим, что область дисперсии $(\Delta\lambda_0)_{\text{о.д}}$ приблизительно в $\mathcal{F}$ раз больше наименьшей разрешаемой разности длин волн. С практически достижимыми величинами $\mathcal{F}$ область дисперсии при большой разрешающей силе очень мала. В типичном случае, рассмотренном выше ($\mathcal{F}=30$, $n'h=4$ *мм*, $\lambda_0=5000$ Å) имеем $(\Delta\lambda_0)_{\text{о.д}}\approx 0{,}3$ Å. В п. 7.6.8 мы увидим, что бóльшую область дисперсии можно получить, пользуясь двумя последовательно соединенными интерферометрами Фабри — Перо. Тем не менее при исследовании сложных спектров во всех случаях (кроме очень близких линий) требуется дополнительная аппаратура для разделения интерференционных картин; это будет описано ниже.

Если две близкие линии разрешены, соответствующий им интервал спектроскопических волновых чисел легко определить из данных о диаметрах колец и расстоянии между пластинами. Согласно (29) и (30) для спектроскопических волновых чисел κ_0 и κ_0' можно написать

$$m_1+e_1=2n'h\kappa_0+\varphi/\pi,\qquad m_1'+e_1'=2n'h\kappa_0'+\varphi/\pi, \tag{48}$$

где m_1 и m_1'— целые порядки первых светлых полос, e_1 и e_1' — дробные порядки в центре картины, и мы предполагаем, что φ постоянно в интервале от κ_0 до κ_0'. Вычитая, получим

$$\kappa_0'-\kappa_0=\frac{(m_1'-m)(e_1'-e_1)}{2n'h}. \tag{49}$$

Для каждой линии дробный порядок получается из измерений диаметров двух светлых колец. Так, из (32) находим, что диаметры D_p и D_q p-го и q-го колец связаны друг с другом соотношением

$$\frac{D_p^2}{D_q^2}=\frac{p-1+e}{q-1+e},$$

откуда

$$e=\frac{(q-1)D_p^2-(p-1)D_q^2}{D_q^2-D_p^2}. \tag{50}$$

Если измерения сделаны не с двумя, а с большим числом колец, то среднюю величину e находят методом наименьших квадратов. Разность целого числа порядков $(m_1'-m_1)$ перекрывающихся картин определяется из наблюдений за изменением картин при уменьшении расстояния между пластинами. Обычно точность в определении величины $(m_1'-m_1)+(e_1'-e_1)$ не превышает $1\cdot 10^{-3}$; так как редко работают с расстоянием между пластинами, меньшем 1 *мм*, то для измерения h достаточно точности микрометра, а n' воздуха можно принять за единицу.

Мы уже говорили, что вследствие малости области дисперсии интерферометра все картины, за исключением картин от очень близких линий, следует разделить. При фотографическом методе регистрации такое разделение обычно успешно достигается «скрещиванием» интерферометра с дифракционным или призменным спектрографом, исправленным на астигматизм. Одно из таких устройств показано на рис. 7.63. Интерферометр I освещается светом источника

S, находящегося в фокальной плоскости линзы L_1, и интерференционная картина проектируется хорошо коррегированной линзой *) L_2 на плоскость щели спектрографа. Интерферометр устанавливают таким образом, чтобы центр картины совпал с центром щели. Тогда щель выделит диаметральное сечение из системы колец, создаваемой каждой присутствующей линией, и задача спектрографа состоит в том, чтобы разделить эти сечения. Если исследуемый спектр состоит из эмиссионных линий, то щель может быть относительно широкой. Тогда в фокальной плоскости спектрографа видны изображения щели в свете каждой линии, пересеченные короткими дугами светлых полос (рис. 7.64). Схожая картина наблюдается и с абсорбционными линиями, если непрерывный фон, сопровождающий линии, захватывает область длин волн, меньшую области дисперсии интерферометра. Абсорбционные линии представляются темными линиями, пересеченными широкими светлыми полосами фона (рис. 7.65). Если непрерывный фон захватывает широкую область длин волн, как,

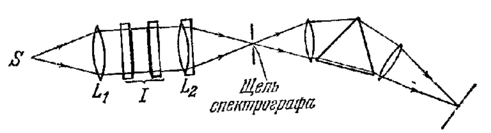


Рис. 7.63 Интерферометр Фабри — Перо, скрещенный с призменным спектрографом.

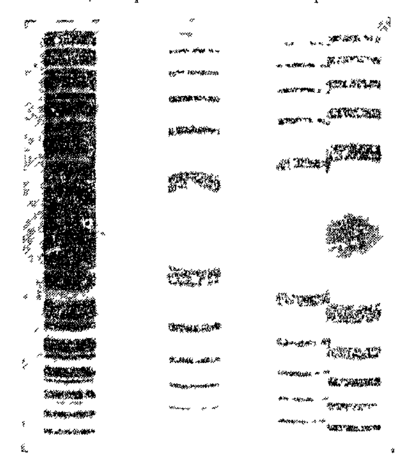


Рис. 7.64. Полосы, полученные с интерферометром Фабри — Перо, в свете эмиссионных линий гелия (фотографический негатив) [70].



Рис. 7.65. Полосы, полученные с интерферометром Фабри — Перо, в свете двух линий спектра неона (фотографический негатив) [71].
а эмиссия б — абсорбция в центре линии с длиной волны 6402 Å показана стрелками

например, в солнечном спектре, ситуация становится более сложной**). Чтобы разобраться в ней, предположим, что ширина щели спектрографа пренебрежимо мала, а разрешающая сила спектрографа бесконечна. Пусть оси прямоугольной системы координат Ox и Oy находятся в фокальной плоскости спектрографа, начало координат O лежит на линии, определяемой светом, прошедшим

*) Возможны и другие комбинации интерферометра со спектрографом. Выбор того или иного устройства в каждом частном случае зависит от таких соображений, как компактность установки, ее стоимость, размеры, интенсивность и однородность источника, а также от присутствия «духов», которые возникают при нежелательных отражениях от поверхностей системы (см [61]).

**) Впервые этот вопрос был разобран Фабри и Бюиссоном [62].

через центр щели, ось Oy параллельна щели. Вдоль линии $x = \text{const}$ длина волны λ_0 постоянна, и положение максимумов интенсивности определяется (28), тогда можно написать

$$\left(m-\frac{\varphi}{\pi}\right)\lambda_0 = 2n'h\cos\theta' = 2n'h\left(1-\frac{\theta'^2}{2}\right) = 2h\left(1-\frac{\theta^2}{2}\right), \quad m = 1, 2, \ldots \quad (51)$$

Мы здесь предполагаем, что θ' мало и принимаем $n' = n = 1$ для воздуха. Угол θ — угол выхода лучей из интерферометра. При малом θ он связан с координатой y соотношением $y = Mf\theta$, где f — фокусное расстояние линзы L_2, проектирующей полосы на щель, а M — увеличение спектрографа. Далее, λ_0 связано с координатой x зависимостью $F(x) = \lambda_0$, где функция F характеризует дисперсию спектрографа. Исключая с помощью этих соотношений λ_0 и θ из (51), мы получим для источника с непрерывным спектром геометрическое место точек максимумов интенсивности

$$\left(m-\frac{\varphi}{\pi}\right)F(x) = 2h\left(1-\frac{y^2}{2M^2f^2}\right),$$
$$m = 1, 2, \ldots \quad (52)$$

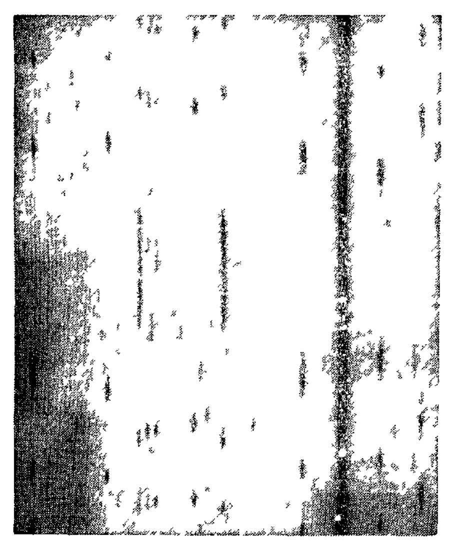


Рис. 7.66. Полосы, полученные с интерферометром Фабри — Перо, обусловленные линиями поглощения в солнечном спектре [72].

Относительное распределение интенсивности между максимумами служит характеристикой интерферометра. Спектр в фокальной плоскости спектрографа рассекается узкими светлыми полосами — каналами с широкими темными промежутками между ними. Эти каналы симметричны относительно линии $y = 0$ и обращены выпуклостью в сторону больших длин волн, при $y = \text{const}$ различие между ними в длинах волн, очевидно, равно области дисперсии интерферометра. В специальном случае, когда $F(x)$ — линейная функция x, каналы имеют вид парабол.

Если в спектре присутствуют линии поглощения, светлые каналы прерываются темными полосами в местах пересечения с линиями поглощения иными словами, эти темные полосы находятся на тех же местах, где были бы светлые полосы, образованные эмиссионными линиями с той же длиной волны, что и у линии поглощения. Если щель раскрывается симметрично, темные полосы расширяются в направлении x. То же происходит и со светлыми полосами, которые в конце концов соприкасаются и сливаются. В этих условиях непрерывный спектр представляется испещренным темными «полосами», обусловленными абсорбционными линиями (рис. 7.66). Полученные картины могут измеряться таким же способом, как и светлые картины, создаваемые эмиссионными линиями. Практически разрешающая сила спектрографа не бесконечна вследствие дифракции и, следовательно, свет с любой длиной волны растягивается на конечное расстояние в направлении x. Таким образом, влияние этого эффекта аналогично расширению щели, т. е. ведет к уширению светлых каналов в направлении x, и они начнут перекрываться, когда спектрограф уже не сможет полностью разрешить длины волн, расстояние между которыми равно расстоя-

нию между соседними светлыми полосами. Мы видели, что это расстояние равно области дисперсии интерферометра, которая уменьшается с увеличением расстояния между пластинами. Тем самым конечная величина разрешающей силы вспомогательного спектрографа ставит верхний предел расстоянию между пластинами, которым еще можно пользоваться.

Интерферометр Фабри — Перо можно также использовать в спектроскопии в комбинации с фотоэлектрическим детектором [63]. Свет исследуемой спектральной линии, выделенный предварительно монохроматором, направляется в интерферометр. Получающаяся интерференционная картина проектируется на кольцеобразное отверстие, концентричное кольцам картины. Это отверстие пропускает свет от небольшой части порядка (кольца) на фотоэлемент. Изменяя оптическое расстояние между пластинами, можно увеличивать или уменьшать величину колец на отверстии и таким образом исследовать структуру интерференционной картины. Такое устройство имеет важное практическое значение потому, что интерферометр Фабри — Перо, как показал Жакино [64], пропускает значительно больший световой поток, чем призменный или дифракционный монохроматор с той же разрешающей силой.

7.6.4. Применение интерферометра Фабри — Перо для сравнения длин волн. Измерения длин волн в спектрах, полученных с призменными или дифракционными спектрометрами, выполняются путем интерполяции, и длина волны любой линии определяется в зависимости от длин волн известных соседних эталонных линий. Так как с большим спектрографом такое сопоставление можно сделать с точностью около $1\cdot 10^{-6}$, то необходимо знать относительные длины волн эталонных линий по крайней мере с такой же точностью Число таких линий также должно быть достаточно велико, чтобы не затруднять интерполяцию. Далее желательно сопоставить длины волн с длиной метра. Таким образом, выбор рациональной системы длин волн для спектроскопических целей включает две операции; (а) сравнение длины волны выбранной первичной эталонной линии с длиной материального эталона метра и (б) сравнение длины волны первичной эталонной линии с длинами волн линий, которые будут служить вторичными эталонными линиями, распределенными по всему спектру. В 1907 г. в качестве первичного эталона [65] была выбрана красная линия (6438 Å) кадмия, возбуждаемая в определенных условиях [66]. Измерение длины волны этой линии в долях метра, является метрологическим процессом, требующим специальных методов, и оно будет описано в § 7.7. Измеренная длина волны, составляющая $6438{,}4696\cdot 10^{-10}$ *м*, принята за 6438,4696 Å. Это и есть определение ангстрема; он отличается от 10^{-10} *м* не больше чем на $3\cdot 10^{-5}$ %. С этим первичным эталоном длин волн чисто оптическим методом сравниваются длины волн всех других линий. Такие измерения составляют важнейшее применение интерферометра Фабри — Перо.

Интерферометр освещают (предпочтительно одновременно) светом первичной эталонной линии и светом неизвестной линии. Интерференционные картины разделяют и фотографируют соответствующим устройством. Пусть λ_s— длина волны эталонной линии в воздухе с показателем преломления среды между пластинами интерферометра, равным n', и пусть λ_1, λ_2, ... — длины волн неизвестных линий. В таком случае, пренебрегая изменением $\varphi\lambda$ с длиной волны, получим из (29) и (30)

$$(m_{1s}+e_s)\lambda_s=(m_{11}+e_1)\lambda_1=(m_{12}+e_2)\lambda_2=\ldots=2H, \tag{53а}$$

$$2H=2h+\frac{\varphi\lambda}{\pi}, \tag{53б}$$

где m_{1s}, m_{11}, m_{12}, ... — целые порядки первых светлых колец, а e_s, e_1, e_2, ... — дробные порядки в центре картины. Для каждой линии дробные порядки можно найти из измерений диаметров колец (см. стр. 309); целые порядки m_{1s}, m_{11}, m_{12}, ... находят методом дробных частей порядка, о котором мы

коротко упоминали на стр. 271. Для этого вначале с помощью микрометра измеряют расстояние h между пластинами. Целое m'_{1s}, ближайшее к $2h/\lambda_s$, является приближенной величиной m_{1s}, и можно написать

$$m_{1s} = m'_{1s} + x, \tag{54}$$

где x — неизвестное целое число. Если неопределенность измерения h составляет Δh, то

$$|x| \leqslant \frac{2\Delta h}{\lambda_s}. \tag{55}$$

Например, при $\Delta h \sim 0{,}01$ *мм* и $\lambda_s = 6438$ Å, имеем $|x| \leqslant 30$. Далее по известным приближенным значениям λ'_1, λ'_2, ... длин волн λ_1, λ_2, ... из (53а) вычисляют приближенные порядки интерференции для неизвестных линий, соответствующие порядку $(m_{1s} + e_s)$ для λ_s; имеем

$$\left.\begin{aligned} m'_{11} + e'_1 &= (m'_{1s} + e_s)\frac{\lambda_s}{\lambda'_1}, \\ m'_{12} + e'_2 &= (m'_{1s} + e_s)\frac{\lambda_s}{\lambda'_2}, \ \dots \end{aligned}\right\} \tag{56}$$

Если принять $\lambda_1 = \lambda'_1 + \Delta\lambda_1$, то из (53а) и (54) находим

$$(m_{11} + e_1)(\lambda'_1 + \Delta\lambda) = (m_{1s} + e_s)\lambda_s = (m'_{1s} + x + e_s)\lambda_s, \tag{57}$$

откуда, используя (56),

$$m_{11} + e_1 = m'_{11} + x + \left[e'_1 + \left(\frac{\lambda_s - \lambda'_1}{\lambda'_1}\right)x\right] - (m_{11} + e_1)\frac{\Delta\lambda'_1}{\lambda'_1}. \tag{58}$$

Дробная часть, содержащаяся в квадратных скобках правой части (58), вычисляется далее для возможных величин x, допускаемых (55). При правильном значении x вычисленная дробь должна согласовываться с измеренной дробью e_1 в левой части (58) с неопределенностью $(m_{11} + e_1)\Delta\lambda'_1/\lambda'_1$, которая мала по сравнению с единицей, если расстояние между пластинами достаточно мало. Например, для приближенных значений длин волн, определенных по измерениям с решеткой, $\Delta\lambda'_1/\lambda'_1 \sim 10^{-5}$; следовательно, $(m_{11} + e_1)\,\Delta\lambda'_1/\lambda'_1 \approx 0{,}1$ для $m_{11} \sim 10^4$, что соответствует $h \approx m_{11}\lambda_1/2 \approx 3$ *мм* при работе в видимой области спектра. Сравнение вычисленных и измеренных дробных частей укажет на непригодность некоторых величин x. Такие же вычисления с приближенными длинами волн λ'_2, λ'_3, ... дают возможность определить x однозначно; обычно достаточно трех линий, если их длины волн расположены соответствующим образом; четвертая линия может служить контрольной.

По известному x определяют m_{1s}, m_{11}, m_{12}, ... из (54) и (58), а из (53а) получим

$$\lambda_1 = \frac{m_{1s} + e_s}{m_{11} + e_1}\lambda_s, \qquad \lambda_2 = \frac{m_{1s} + e_s}{m_{12} + e_2}\lambda_s, \ \dots \tag{59}$$

Так как целые порядки определяются точно, то величины $\Delta\lambda_1$, $\Delta\lambda_2$, ... для этих длин волн обусловлены только ошибками в измерении дробей e_s, e_1, e_2, ..., и если последние не превышают $\pm 0{,}01$ — точность, которую нетрудно достичь с высоко отражающими покрытиями, а m_{1s}, m_{11}, m_{12}, ... порядка 10^4, как в только что приведенном примере, то $\Delta\lambda_1/\lambda_1$, $\Delta\lambda_2/\lambda_2$, ... будут порядка 10^{-6}, т. е. при таких измерениях точность, с которой известны λ_1, λ_2, ... увеличилась примерно в 10 раз. Исправленные значения длин волн можно использовать для повторных опытов с еще бо́льшим расстоянием между пластинами, что в свою очередь дает еще бо́льшую точность для значений длин волн. Подобную процедуру можно повторять до тех пор, пока резкость линий оправдывает эти измерения.

Выше мы предполагали, что (не считая первичных эталонов) имеются лишь относительно неточные значения длин волн *). В свое время было измерено большое число вторичных эталонов с точностью, близкой к точности определения первичных эталонов, что значительно упростило дальнейшие измерения длин волн. Эталонные длины волн, а также метод дробных частей порядка **) используются для определения $2H$. Для однозначного определения по приближенным значениям длин волн λ_1', λ_2', ... величин m_{11}, m_{12}, ... нужно только, чтобы величина $2H$ была достаточно мала. Так, из (53а) и (57) имеем

$$m_{11}+e_1=\frac{2H}{\lambda_1'}-(m_{11}+e_1)\frac{\Delta\lambda_1'}{\lambda_1'}, \tag{60}$$

а m_1 получается однозначно, если, например, $(m_{11}+e_1)\,\Delta\lambda_1'/\lambda_1'\approx 0{,}3$. Пусть, как и раньше, $\Delta\lambda_1'/\lambda_1'\sim 10^{-5}$, тогда $m_{11}\approx 3\cdot 10^4$, что соответствует $H\approx 10$ *мм* для длин волн видимого света; если измеряемая дробь e_1 лежит в пределах $\pm 0{,}01$, погрешность $\Delta\lambda_1/\lambda_1$ в длине волны $\lambda_1=2H/(m_{11}+e_1)$ примерно равна $3\cdot 10^{-7}$, т. е. однократное применение интерферометра увеличивает точность почти в тридцать раз.

Практически предположение о независимости $\varphi\lambda$ от длины волны может оказаться неверным, если сравниваемые линии широко разнесены по спектру. Систематические ошибки, связанные с этим, устраняются, если измерения производятся с двумя значениями h, например с h_{I} и h_{II}. Тогда, применяя те же обозначения, что и в (53), находим из (29) и (30)

$$\left\{(m_{1s}+e_s)_{\text{I}}-\frac{\varphi_s}{\pi}\right\}\lambda_s=\left\{(m_{11}+e_1)_{\text{I}}-\frac{\varphi_1}{\pi}\right\}\lambda_1=\ldots=2h_{\text{I}},$$

$$\left\{(m_{1s}+e_s)_{\text{II}}-\frac{\varphi_s}{\pi}\right\}\lambda_s=\left\{(m_{11}+e_1)_{\text{II}}-\frac{\varphi_1}{\pi}\right\}\lambda_1=\ldots=2h_{\text{II}},$$

откуда

$$\lambda_1=\frac{(m_{1s}+e_s)_{\text{II}}-(m_{1s}+e_s)_{\text{I}}}{(m_{11}+e_1)_{\text{II}}-(m_{11}+e_1)_{\text{I}}}\lambda_s. \tag{61}$$

Длины волн, полученные после коррекции на изменение фазы, приводят к их величинам в нормальных условиях, при которых измерялся первичный эталон (сухой воздух при 15° С и давлении 760 *мм рт. ст.*). Для этого необходимо знать дисперсию воздуха. Если далее следует привести найденные длины волн к их значениям в вакууме, то нужно знать еще и показатель преломления воздуха. Величины показателя преломления и дисперсии воздуха с точностью, вполне достаточной для данной цели, можно получить с помощью многолучевых интерферометрических методов (см. п. 7.6.8).

7.6.5. **Интерферометр Люммера — Герке.** Мы уже упоминали в п. 7.6.1, что коэффициент отражения света на внутренней поверхности диэлектрика приближается к единице, когда угол падения в более плотной среде приближается к критическому углу. Это обстоятельство использовано в многолучевом интерферометре, предложенном Люммером и Герке [73], основная часть которого представляет собой длинную плоскопараллельную пластинку из стекла или кристаллического кварца (рис. 7.67). Пучок света от источника, лежащего

*) В своих ранних работах по сравнению длин волн с длиной волны красной линии кадмия Фабри и Перо [67] пользовались интерферометром с подвижным зеркалом. Они использовали точные длины волн зеленой (5086 Å) и голубой (4800 Å) линий кадмия, измеренные на двухлучевом интерферометре [26], и определяли целые порядки с помощью так называемого *«метода совпадений»*. Принципиально этот метод не отличается от описанного выше метода дробных частей порядка, но более трудоемок и требует визуальных наблюдений.

**) Если эталонные длины волн расположены достаточно часто, то можно быстро подсчитать приближенную величину x и тем самым сократить вычисления (см., например, [68]). Следует упомянуть, что Гаррисон [69] описал машину, с помощью которой длины волн определяются непосредственно по интерференционной картине, и все необходимые вычисления производятся автоматически.

на продольной оси пластинки, входит в нее через призму P, укрепленную на одном из концов пластинки, и падает на внутреннюю поверхность последней под углом, немного меньшим критического. В этом случае ряд световых пучков выходит с каждой стороны пластинки под углом, очень близким к скользящему; вышедшие из нее пучки собираются линзой L и образуют интерференционную картину в ее фокальной плоскости.

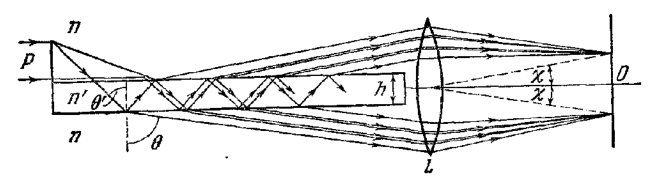


Рис. 7.67. Пластинка Люммера — Герке.

Для монохроматического света с длиной волны λ_0 разность фаз δ между соседними пучками определяется выражением (1), т. е.

$$\delta = \frac{4\pi}{\lambda_0} n'h\cos\theta' = \frac{4\pi h}{\lambda_0}\sqrt{n'^2 - n^2\sin^2\theta}, \tag{62}$$

где h и n' — толщина пластинки и ее показатель преломления, n — показатель преломления окружающего воздуха, θ' — угол отражения в пластинке, а θ — угол выхода лучей из пластинки. Соответствующий порядок интерференции m равен

$$m = \frac{\delta}{2\pi} = \frac{2h}{\lambda_0}\sqrt{n'^2 - n^2\sin^2\theta}. \tag{63}$$

Полосы, соответствующие постоянному θ, образуют семейство гипербол, которые вблизи центра O картины имеют вид почти прямых линий, параллельных поверхностям пластинки. В соответствии с (63) для порядка интерференции m_0 в центре ($\theta = \pi/2$) получим

$$m_0 = \frac{2h}{\lambda_0}\sqrt{n'^2 - n^2}, \tag{64}$$

и

$$m - m_0 = \frac{2h}{\lambda_0}\left(\sqrt{n'^2 - n^2\sin^2\theta} - \sqrt{n'^2 - n^2}\right). \tag{65}$$

Если мы положим $\chi = (\pi/2) - \theta$, то χ мало́ вблизи центра, и можно считать, что $\sin^2\theta = 1 - \chi^2$. Тогда, разлагая в ряд (65) и отбрасывая члены с χ в степени, большей второй, находим

$$m - m_0 = \frac{n^2}{\sqrt{n'^2 - n^2}}\,\frac{h\chi^2}{\lambda_0}. \tag{66}$$

Угол χ_q, соответствующий q-й от центра светлой полосе целого порядка m_q, определяется выражением

$$\chi_q = \frac{\sqrt[4]{n'^2 - n^2}}{n}\sqrt{\frac{\lambda_0}{h}}\sqrt{m_q - m_0} = \frac{\sqrt[4]{n'^2 - n^2}}{n}\sqrt{\frac{\lambda_0}{h}}\sqrt{q - e}, \tag{67}$$

где e — дробный порядок в центре. Следовательно, угловой масштаб картины пропорционален $\sqrt{\lambda_0/h}$, и если $e = 0$, то расстояния светлых полос от оси пропорциональны квадратному корню из целых положительных чисел.

Распределение интенсивности в такой картине отличается от распределения, рассмотренного в п. 7.6.1, так как в данном случае отсутствует вклад света, вносимый первым отражением поверхностью пластины. Чтобы упростить здесь обсуждение, рассмотрим схему, показанную на рис. 7.68. Свет попадает в пластинку через отверстие JK, ограниченное диафрагмой $\mathcal{AB}$, и преломляется

в пластинке. Очевидно, что пластинка, соединенная с призмой, изображенной на рис. 7.67, действует так же, как и пластинка с отверстием. Разница состоит только в том, что у пластинки с призмой коэффициент отражения для входящего луча ничтожно мал *). Пусть P и P' — фокальные точки L для света, выходящего из нижней и верхней поверхностей пластинки под углом 0, и пусть A — комплексная амплитуда света в P в отсутствие пластинки. Комплексные

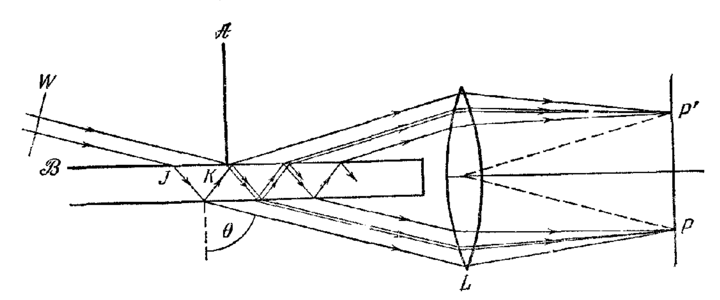


Рис. 7.68. Многократное отражение в плоскопараллельной пластинке при устранении отражений первичного пучка света от ее внешней поверхности.

амплитуды в P в пучках нижней группы равны (см. п. 7.6.1), с точностью до несущественного постоянного фазового множителя, величинам

$$tt'A,\ tt'r'^2A\exp(i\delta),\ \ldots,\ tt'r'^{(2p-1)}A\exp(i(p-1)\delta),$$

и, следовательно, для результирующей амплитуды в точке P при суперпозиции p первых пучков имеем

$$A^{(t)}(p)=(1+r'^2\exp(i\delta)+\ldots+r'^{(2p-1)}\exp[i(p-1)\delta]\,tt'A=$$
$$=\frac{1-\mathcal{R}^p\exp(ip\delta)}{1-\mathcal{R}\exp(i\delta)}\mathcal{T}A. \quad (68)$$

В пределе, считая, что предельное число пучков $p\to\infty$, находим

$$A^{(t)}\equiv A^{(t)}(\infty)=\frac{\mathcal{T}}{1-\mathcal{R}\exp(i\delta)}A, \quad (69)$$

а интенсивность $I^{(t)}=A^{(t)}A^{(t)*}$ в P равна

$$I^{(t)}\equiv I^{(t)}(\infty)=\frac{\mathcal{T}^2}{1+\mathcal{R}^2-2\mathcal{R}\cos\delta}I=\frac{1}{1+F\sin^2\frac{\delta}{2}}I, \quad (70)$$

где $I=AA^*$ — интенсивность в P в отсутствие пластинки. Аналогично для амплитуд в точке P' в пучках от верхней поверхности имеем

$$tt'r'A,\ tt'(r')^3\exp(i\delta)A,\ \ldots\ tt'r'^{(2p-1)}\exp[i(p-1)\delta]A\ldots,$$

и, следовательно, при суперпозиции p первых пучков результирующая амплитуда в точке P' равна

$$A^{(r)}(p)=(1+r'^2\exp(i\delta)+\ldots+r'^{2(p-1)}\exp[i(p-1)\delta]\,tt'r'A=$$
$$=\frac{1-R^p\exp(ip\delta)}{1-R\exp(i\delta)}\sqrt{\mathcal{R}}\,\mathcal{T}A, \quad (71)$$

и в пределе, когда $p\to\infty$, интенсивность в P' запишется в виде

$$I^{(r)}\equiv I^{(r)}(\infty)=\frac{\mathcal{R}\mathcal{T}^2}{1+\mathcal{R}^2-2\mathcal{R}\cos\delta}I=\mathcal{R}I^{(t)}. \quad (72)$$

*) Подобное устройство реализуется на практике, если свет поступает в интерферометр Фабри — Перо через прозрачное отверстие в одном из отражающих покрытий. Особенности такого устройства рассмотрены Дюфуром [74].

Из (70) и (72) очевидно, что максимумы в интерференционных картинах, лежащих под и над пластинкой, находятся на одинаковых угловых расстояниях относительно нормали к ней.

Выходящие друг за другом с одной стороны пластинки пучки разделены, очевидно, расстоянием $2h\,\mathrm{tg}\,\theta'$ (считая вдоль пластинки) и при пластинке длиной l общее число пучков p равно

$$p \approx \frac{l}{2h}\,\mathrm{ctg}\,\theta' = \frac{l}{2h}\sqrt{\frac{n'^2}{n^2\sin^2\theta} - 1}, \tag{73}$$

или для $\theta \approx \pi/2$ и при $n=1$ для воздуха

$$p \approx \frac{l}{2h}\sqrt{n'^2 - 1}. \tag{74}$$

При приближении θ' к критическому углу $\mathcal{R} \to 1$, а $\mathcal{R}^p$ перестает быть ничтожно малым, и поэтому, использование предельных выражений (70) и (72) с $p \to \infty$ становится неоправданным; в этих условиях распределение интенсивности в верхней картине получается из (68) в виде

$$I^{(t)}(p) = \frac{1 + \mathcal{R}^{2p} - 2\mathcal{R}^p\cos(p\delta)}{1 + \mathcal{R}^2 - 2\mathcal{R}\cos\delta}\,\mathcal{T}^2 I =$$

$$= \frac{(1-\mathcal{R}^p)^2 + 4\mathcal{R}^p\sin^2\left(p\frac{\delta}{2}\right)}{(1-\mathcal{R})^2 + 4\mathcal{R}\sin^2\left(\frac{\delta}{2}\right)}\,\mathcal{T}^2 I = \frac{1 + G_p\sin^2\left(\frac{p\delta}{2}\right)}{1 + F\sin^2\left(\frac{\delta}{2}\right)}\,(1-\mathcal{R}^p)^2 I, \tag{75}$$

где

$$G_p = \frac{4\mathcal{R}^p}{(1-\mathcal{R}^p)^2}, \tag{76}$$

а F задано (16). Итак, распределение интенсивности определяется функцией $(1 + G_p\sin^2(p\delta/2))/(1 + F\sin^2(\delta/2))$, отличающейся от функции распределения при $p \to \infty$ добавочным членом $(G_p\sin^2(p\delta/2))/(1 + F\sin^2(\delta/2))$. Последний не влияет на положение абсолютных максимумов, остающихся при $\delta/2 = m\pi$, где m — целое число, но так как $(G_p\sin^2(p\delta/2))/(1 + F\sin^2(\delta/2))$ равно нулю при $\delta/2 = (m + q/p)\pi$, $q = 0, 1, 2, \ldots, p$ и положительно в остальных случаях, то абсолютные максимумы становятся шире, чем при $p \to \infty$, и между ними появляются вторичные максимумы (рис. 7.69).

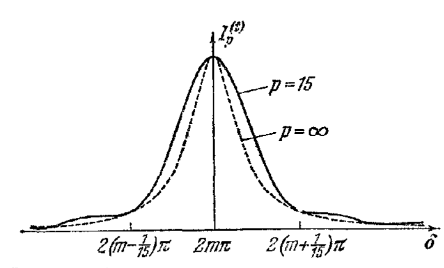


Рис 7.69. Зависимости интенсивности в интерференционных полосах, полученных с пластинкой Люммера — Герке, от разности фаз ($\mathcal{R}$ 0,87)

Пластинка Люммера — Герке служит исключительно для исследования тонкой структуры спектральных линий. Применяется она, подобно интерферометру Фабри — Перо, в комбинации со вспомогательным диспергирующим прибором. Область дисперсии пластинки можно найти из формулы (63), которая, если считать $n=1$, запишется в виде

$$m^2\lambda_0^2 = 4h^2(n'^2 - \sin^2\theta). \tag{77}$$

Таким образом, в данном угловом положении изменение порядка Δm, соответствующее изменению длины волны на $\Delta\lambda_0$, с точностью до членов первого порядка относительно малых величин равно

$$\Delta m = \frac{4h^2 n'\frac{\partial n'}{\partial\lambda_0} - m^2\lambda_0}{m\lambda_0^2}\,\Delta\lambda_0. \tag{78}$$

Для области дисперсии, соответствующей изменению m на единицу, получим, следовательно,

$$(\Delta\lambda_0)_{\text{о. д}} = \frac{m\lambda_0^2}{\left| m^2\lambda_0 - 4h^2n' \frac{\partial n'}{\partial\lambda_0} \right|}, \tag{79}$$

или, подставляя m из (77) и полагая $\sin\theta \sim 1$ для лучей, выходящих под углом, близким к скользящему,

$$(\Delta\lambda_0)_{\text{о. д}} \approx \frac{\lambda_0^2}{2h} \frac{\sqrt{n'^2-1}}{\left| n'^2 - n'\lambda_0 \frac{\partial n'}{\partial\lambda_0} - 1 \right|}. \tag{80}$$

Мы видим, что величина $(\Delta\lambda_0)_{\text{о. д}}$ обратно пропорциональна толщине пластинки. У кварца множитель $\left[\sqrt{n'^2-1} \Big/ \left(n'^2 - n'\lambda_0 \frac{\partial n'}{\partial\lambda_0} - 1\right)\right]$ изменяется приблизительно от 0,6 при $\lambda_0 = 2000$ Å до 0,8 при $\lambda_0 = 6000$ Å.

Если считать $\mathscr{R}$ постоянной, то точки в интерференционной картине, где $I^{(t)}(p)$ уменьшается до половины своей максимальной величины, соответствуют $\delta = 2m\pi \pm \varepsilon/2$ (m целое число), и из (75) находим

$$\frac{1+G_p \sin^2\left(\frac{p\varepsilon}{4}\right)}{1+F\sin^2\left(\frac{\varepsilon}{4}\right)} = \frac{1}{2},$$

т. е.

$$\sin^2\left(\frac{p\varepsilon}{4}\right) - \frac{F}{2G_p}\sin\left(\frac{\varepsilon}{4}\right) + \frac{1}{2G_p} = 0. \tag{81}$$

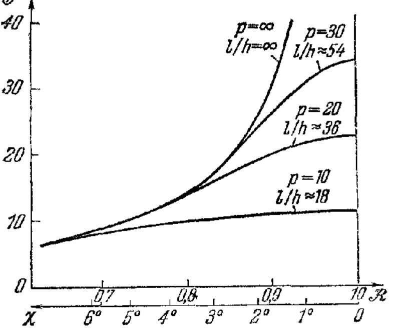


Рис. 7.70. Зависимость резкости $\mathscr{F}$ от отражательной способности $\mathscr{R}$ и числа отражений p для пластинки Люммера — Герке. Для сравнения указаны величины угла χ, отсчитываемого от центра картины, и отношение длины кварцевой пластинки к ее толщине l/h.

Очевидно, что резкость $\mathscr{F} = 2\pi/\varepsilon$ зависит как от $\mathscr{R}$, так и от p. На рис. 7.70 показана зависимость ее от $\mathscr{R}$ для различных значений p. Согласно (74) p определяется в основном отношением l/h. $\mathscr{R}$ является, конечно, функцией θ', а значит, и χ; следовательно, при данном χ значение $\mathscr{R}_\perp$ больше $\mathscr{R}_\parallel$ (при $n' > n$); поэтому выгодно выделять поляризованную компоненту, колеблющуюся перпендикулярно к плоскости падения. Предельный случай $p \to \infty$ соответствует пластинке бесконечной длины, у которой распределение интенсивности такое же, как и в интерферометре Фабри — Перо (см. (70) и (27)). В этом случае резкость определяется из (35). Рис. 7.70 показывает, что при конечном p величина $\mathscr{F}$ соответствует $p = \infty$, пока $\mathscr{R}$ не слишком велико. При таких значениях $\mathscr{R}$ можно пренебречь $\mathscr{R}^p$ (а значит, и G_p) и тогда, по существу, весь свет, вошедший в пластинку, участвует в образовании полос. При больших значениях $\mathscr{R}$ величина $\mathscr{F}$ меньше величины, соответствующей $p = \infty$, и когда $\mathscr{R}$ приближается к единице, $\mathscr{F}$ приближается к максимальной величине

$$\mathscr{F}_l = \frac{2\pi}{\varepsilon_l}, \tag{82}$$

где предельную величину ε_l можно найти из (81). Например, полагая $\mathscr{R} = 1 - \mathscr{T}$, получим (допуская, что $\mathscr{T}$ мало) $1 - \mathscr{R}^p \approx p\mathscr{T}$, так что $F/G_p = (1-\mathscr{R}^p)^2/\mathscr{R}^{p-1}(1-\mathscr{R})^2 \approx p^2/(1-\mathscr{T})^{p-1} \to p^2$ при $\mathscr{T} \to 0$. Поскольку $1/G_p \to 0$, когда $\mathscr{T} \to 0$, (81) дает

$$\sin^2\left(\frac{p\varepsilon_l}{4}\right) - \frac{p^2}{2}\sin^2\left(\frac{\varepsilon_l}{4}\right) = 0. \tag{83}$$

Отсюда для малого ε_l находим

$$\frac{\sin(p\varepsilon_l/4)}{p\varepsilon_l/4}=\frac{1}{\sqrt{2}}, \qquad \frac{p\varepsilon_l}{4}=0,45\pi, \quad \mathcal{F}_l=1,1p. \tag{84}$$

Однако из (75) следует, что при $\mathcal{R}\to 1$ интенсивность $I^{(t)}(p)$ в интерференционной картине стремится к нулю, ибо увеличивается доля света, который, войдя в пластинку, остается в ней после последнего отражения и теряется у ее дальнего конца. Если, как обычно, любые заметные потери света недопустимы, пластинка должна работать в той области, где величина $\mathcal{F}$ соответствует $p=\infty$, и из рис. 7.70 следует, что тогда наибольшее значение $\mathcal{F}$ приблизительно равно $2\mathcal{F}_l/3\approx 0,7\,p$. В этой области относительное распределение интенсивности такое же, как и в интерферометре Фабри — Перо. Таким образом, как мы видели на стр. 307, можно принять, что наименьшая разрешаемая разность длин волн составляет приблизительно $1/\mathcal{F}$ часть области дисперсии $(\Delta\lambda_0)_{\text{о.д}}$. Из (74) и (80) получим выражение для разрешающей силы, соответствующее $\mathcal{F}\approx 0,7\,p$, в виде

$$\frac{\lambda_0}{\Delta\lambda_0}\approx\frac{0,7p\lambda_0}{(\Delta\lambda_0)_{\text{о. д}}}\approx 0,7\,\frac{l}{\lambda_0}\left|n'^2-n'\lambda_0\frac{\partial n'}{\partial\lambda_0}-1\right|. \tag{85}$$

Разрешающая сила, определяемая последним соотношением, зависит только от длины пластинки и не зависит от ее толщины, но следует помнить, что она достигается при величинах θ, приближающихся к $\pi/2$ при увеличении l. Практически длина l ограничивается техническими трудностями, возникающими при изготовлении пластинки.

Пластинка Люммера — Герке, сделанная из кристаллического кварца, прозрачна для длин волн вплоть до 2000 Å, и в течение долгого времени она была наилучшим из доступных интерферометров с большой разрешающей силой для спектроскопии в ультрафиолетовой области. Однако получение отражающих покрытий для ультрафиолета, пригодных для использования в интерферометре Фабри — Перо, и успехи в изготовлении отражающих решеток типа эшелетов привели к тому, что теперь пластинка Люммера — Герке используется в исследованиях крайне редко.

7.6.6. Интерференционные фильтры. Предположим, что параллельный пучок белого света падает нормально на плоскопараллельную пластинку с поверхностями, хорошо отражающими свет. В соответствии с (27) мы видим, что в таком случае в прошедшем свете появляются максимумы интенсивности, если разность фаз δ равна целому числу, умноженному на 2π, или, согласно (25), если $\lambda_0=\lambda_0^{(m)}$, где

$$\lambda_0^{(m)}=\frac{2n'h}{(m-\varphi/\pi)}, \qquad m=1,\ 2,\ \ldots, \tag{86}$$

и по обе стороны от этих максимумов интенсивность прошедшего света быстро падает до малых величин. Таким образом, пластинка действует как фильтр длин волн с многочисленными полосами пропускания, получающимися при целых значениях порядка m. В частности, если оптическая толщина пластинки составляет лишь несколько полуволн видимого света, то полосы пропускания в видимой области принадлежат низким порядкам и широко разнесены по длинам волн. Поэтому обычно можно не пропустить все эти полосы, кроме одной, либо при помощи вспомогательных абсорбционных фильтров, либо, воспользовавшись селективной реакцией детектора, применяемого для наблюдения.

Фильтр подобного типа изготовляют, нанося на плоскую поверхность стеклянной пластинки две отражающие пленки, разделенные слоем диэлектрика (рис. 7.71). Если при нормальном падении фильтр должен пропускать полосу порядка m длины волны $\lambda_0^{(m)}$, то, согласно (86), оптическая толщина $n'h$ промежуточного слоя должна равняться $(m-\varphi/\pi)\,\lambda_0^{(m)}/2$. У металлических

отражающих пленок φ зависит от λ_0 и толщины пленки (и, следовательно, от отражательной способности $\mathcal{R}$). У многослойных диэлектрических отражающих пленок $\mathcal{R}(\lambda_0^{(m)})$ максимально, когда оптическая толщина каждого слоя равна $\lambda_0^{(m)}/4$; в этом случае $\varphi(\lambda_0^{(m)})$ равно нулю (см. § 1.6) и, следовательно, оптическая толщина промежуточного слоя должна составлять $m\lambda_0^{(m)}/2$.

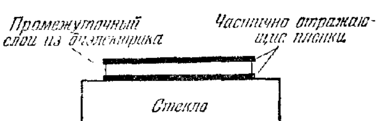


Рис. 7.71. Интерференционный фильтр типа Фабри — Перо.

Важными характеристиками такого фильтра служат *максимальное пропускание* τ, определяемое так же, как и у интерферометра Фабри — Перо (см. (33)), и *полуширина* полосы пропускания $(\Delta\lambda_0)_{п.ш.}$, которая определяется как интервал между длинами волн в полосе пропускания, на котором отношение $I^{(t)}/I^{(i)}$ уменьшается до половины своей максимальной величины. Если оптическое расстояние между отражающими поверхностями постоянно по всей рабочей поверхности фильтра, то τ получается из (36), а согласно (21) $(\Delta\lambda_0)_{п.ш.}$ соответствует изменению δ на величину $4/\sqrt{F}$ при условии, что в области пропускания фильтра мы вправе пренебречь изменениями $\mathcal{R}$ и $\mathcal{F}$ с длиной волны. Итак, если можно пренебречь дисперсией в промежуточном слое, то при нормальном падении света изменению длины волны на $\Delta\lambda_0$ соответствует изменение δ на $\Delta\delta$, равное с точностью до членов первого порядка относительно малых величин

$$\Delta\delta = \left(-\frac{4\pi}{\lambda_0^2}n'h + 2\frac{\partial\varphi}{\partial\lambda_0}\right)\Delta\lambda_0, \tag{87}$$

и вблизи полосы пропускания порядка m мы получим, используя (86),

$$\Delta\delta = \frac{2\pi}{\lambda_0^{(m)}}\left\{-m + \frac{1}{\pi}\left[\frac{\partial}{\partial\lambda_0}(\varphi\lambda_0)\right]_{\lambda_0^{(m)}}\right\}\Delta\lambda_0. \tag{88}$$

Поэтому, учитывая (22), находим для полуширины фильтра, выраженной в длинах волн, которая, как мы только что показали, соответствует изменению в δ на $4/\sqrt{F}$;

$$(\Delta\lambda_0)_{п.ш} = \frac{2\lambda_0}{\pi\sqrt{F}\left|m - \frac{1}{\pi}\left[\frac{\partial}{\partial\lambda_0}(\varphi\lambda_0)\right]_{\lambda_0^{(m)}}\right|} = \frac{\lambda_0}{\mathcal{F}\left|m - \frac{1}{\pi}\left(\frac{\partial}{\partial\lambda_0}(\varphi\lambda_0)\right)_{\lambda_0^{(m)}}\right|}. \tag{89}$$

У металлических отражающих пленок $\partial\varphi/\partial\lambda_0$ столь мало, что $\partial(\varphi\lambda_0)/\partial\lambda_0 \approx \varphi$. Дюфур [75] исследовал влияние изменения фазы в том случае, когда в отражающие пленки входит многослойный диэлектрик, и показал, что в видимом спектре у таких диэлектриков, как сульфид цинка и криолит, $\frac{\partial}{\partial\lambda_0}(\varphi\lambda_0)/\pi$ находятся между *) $-1{,}0$ и $-1{,}5$.

Мы видели, что при имеющихся в нашем распоряжении отражающих пленках τ уменьшается при увеличении $\mathcal{F}$; следовательно, для фильтра заданного порядка, изготовленного из данных материалов, меньшей полуширине сопутствует меньшее пропускание. Практически, как и в случае интерферометра Фабри — Перо, верхний предел полезного отражения определяется изменениями оптического расстояния между отражающими поверхностями по рабочей площади пластин. Однако здесь отступление от идеальной плоскости стеклянной поверхности, на которой изготовляется фильтр, несущественно, так как нанесенная испарением пленка повторяет все неровности поверхности, на

*) При исследовании разрешающей силы интерферометра Фабри — Перо мы пренебрегали зависимостью φ от длины волны (см. (42) и (88)). Это, очевидно, оправдано, когда m достаточно велико.

которую она наносится. При соответствующем изготовлении фильтров неравномерность в оптическом расстоянии между отражающими поверхностями удается сделать значительно меньше, чем при двух независимо обрабатываемых рабочих поверхностях в интерферометре Фабри — Перо, и поэтому можно использовать соответственно бо́льшую отражательную способность. Типичные характеристики фильтров с металлическими и комбинированными отражающими

Таблица 7.2

Характеристика интерференционных фильтров с металлическими отражающими пленками и пленками металл-диэлектрик [76]

Тип фильтра	Длина волны $\lambda_0^{(m)}$ максимума пропускания, Å	Максимальное пропускание τ	Полуширина $(\Delta\lambda_0)$ п. ш, Å
M—2L—M	5310	0,30	130
M—4L—M	5350	0,26	70
MLH—2L—HLM	5470	0,43	48
MLHLH—2L—HLHLM	6050	0,38	20

M — металлическая пленка (серебро), *L* — четвертьволновой слой диэлектрика с малым показателем преломления (фтористый магний), *H* — четвертьволновой слой диэлектрика с большим показателем преломления (сульфид цинка).

покрытиями металл — диэлектрик приведены в табл. 7.2, а с чисто диэлектрическими покрытиями — в табл. 7.3. Как мы видим, наибольшая величина пропускания при данной полуширине достигается с отражающими пленками, содержащими слои диэлектрика. Следует заметить, что отражающие свойства нанесенных пленок зависят от длины волны, и поэтому такие фильтры дают вторичные полосы пропускания, помимо тех, которые определяются выражениями (86).

Таблица 7.3

Характеристика интерференционных фильтров из сульфида цинка — криолита [77]

Тип фильтра	Длина волны $\lambda_0^{(m)}$ максимума пропускания, Å	Максимальное пропускание, τ	Полуширина $(\Delta\lambda_0)$ п. ш, Å
HLH—2L—HLH	5185	0,90	380
HLHLH—2L—HLHLH	4750	0,85	110
HLHLHL—2H—LHLHLH	6565	0,90	65
HLHLHLH—2L—HLHLHLH	5200	0,70	40
HLHLHLHL—2H—LHLHLHLH	6500	0,80	35
HLHLHLHLHL—2H—LHLHLHLHLH	6600	0,50	20

L — четвертьволновой слой криолита, *H* — четвертьволновой слой сульфида цинка.

Так как фильтры должны пропускать только определенную длину волны, оптическую толщину промежуточного слоя следует тщательно контролировать. Согласно (86) изменению $n'h$ на $\Delta(n'h)$ соответствует смещение $\Delta\lambda_0^{(m)}$ полосы пропускания m-го порядка на величину $\Delta\lambda_0^{(m)} = \lambda_0^{(m)}\Delta(n'h)/n'h$. Например, в результате ошибки в оптической толщине промежуточного слоя на 1% при $\lambda_0^{(m)} = 5000$ Å полоса пропускания фильтра сместится на 50 Å, что превышает полуширину узкополосных фильтров. Удобные методы контроля толщины про-

межуточного слоя во время его нанесения были описаны Гриндландом и Биллингтоном [78], Джиакомо и Жакино [79], и Лисбергером и Рингом [80]. Из (25) видно также, что полосу пропускания фильтра можно сместить в сторону более коротких волн, наклоняя фильтр так, чтобы падение света не было нормальным. Однако качество фильтра ухудшается по мере увеличения наклона. В частности, при металлических или отражающих пленках фазы $\varphi_{\parallel}$ и $\varphi_{\perp}$ для света, поляризованного параллельно и перпендикулярно плоскости падения, становятся неравными, и полосы пропускания таких, по-разному поляризованных волн будут находиться при разных длинах волн. Смещение полосы пропускания при заданном угле наклона зависит от угла отражения θ' в промежуточном слое таким образом, что меньшее смещение полосы соответствует большему показателю преломления вещества этого слоя. Поэтому промежуточный слой выгодно изготовлять из материалов с большим показателем преломления, в особенности если фильтр должен работать в сходящемся пучке.

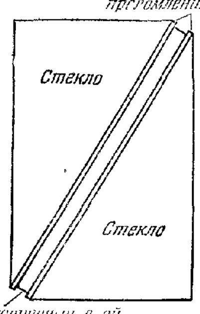


Рис. 7.72. Фильтр с нарушенным полным внутренним отражением.

В другом типе интерференционного фильтра, называемом *фильтром с нарушенным полным внутренним отражением*, впервые описанным в работе [81], тонкие отражающие слои из вещества с малым показателем преломления окружены веществом с большим показателем преломления. В п. 1.5.4 было показано, что при освещении такого достаточно тонкого слоя светом под углом, бóльшим критического, отражение не будет полным, и некоторое количество света пройдет сквозь слой, который в этих условиях действует как непоглощающий отражатель. Изменяя толщину слоя, можно получить любую степень отражения. Конструкция такого фильтра показана на рис. 7.72. У призмы из тяжелого стекла гипотенузную грань покрывают слоем вещества с малым показателем преломления, на который последовательно наносят промежуточный слой с высоким показателем преломления и слой с малым показателем преломления; затем вторую призму, подобную первой, соединяют с первой веществом с таким же показателем преломления, как и у стекла.

Угол призмы и показатели преломления выбирают так, чтобы свет, падающий нормально основанию призмы, попадал на слой с малым показателем преломления под углом, бóльшим критического. Длины волн полос пропускания

Таблица 7.4

Характеристики некоторых фильтров с нарушенным полным внутренним отражением с отражающими слоями из фтористого магния на тяжелом флинте и с промежуточным слоем из сульфида цинка [82]

Порядок интерференции	1		2		3	
Поляризация	$\parallel$	$\perp$	$\parallel$	$\perp$	$\parallel$	$\perp$
Длина волны максимума пропускания $\lambda_0^{(m)}$, Å	4600	5300	4630	5080	—	5110
Максимальное пропускание τ	>0,93	>0,93	0,90	0,90	—	0,12
Полуширина $\lvert\Delta\lambda_0\rvert_{п.ш}$ Å	66	120	50	49	—	30

фильтра зависят от оптической толщины промежуточного слоя, угла преломления в нем и сдвига фаз при отражении. Отражательная способность (и, следовательно, полуширина для данного порядка) зависит от толщины слоев с низким показателем преломления. Сдвиг фазы зависит от состояния поляризации, и поэтому для данного порядка полосы пропускания компонент, колеблющихся параллельно или перпендикулярно плоскости падения, имеют разные длины волн. Характеристики типичных фильтров, работающих в видимом спектре, приведены в табл. 7.4. Фильтры такого типа изготовляются и для работы в сантиметровом диапазоне [83].

7.6.7. Многолучевые полосы в тонких пленках. В п. 7.6.1 мы видели, что полосы равного наклона, создаваемые в плоскопараллельной пластинке, сильно сужаются при увеличении отражательной способности ее поверхностей. Подобное же действие оказывает увеличение отражательной способности на распределение интенсивности в полосах равной толщины Физо (см. п. 7.5.2), наблюдаемых в тонких пленках. В результате такого перераспределения интенсивности отчетливо выступают более тонкие изменения в толщине пленки.

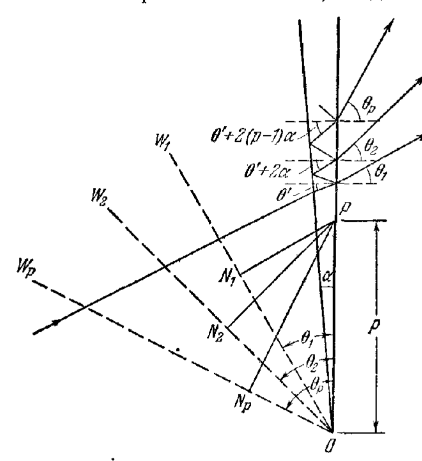


Рис. 7.73. Многократные отражения в клине.

Для простоты рассмотрим сперва пленку в виде клина, образованного плоскими поверхностями с небольшим углом α. На пленку под прямым углом к ребру *) падает плоская волна монохроматического света (рис. 7.73). Вследствие многократных отражений от поверхностей прошедший свет состоит из группы плоских волн, распространяющихся в различных направлениях. Если прошедшая через первую поверхность волна падает на вторую поверхность со стороны нормали, ближайшей к ребру клина, под углом θ', то p-я волна прошедшей группы выходит из клина под углом θ_p, и из законов преломления и отражения следует, что

$$n \sin \theta_p = n' \sin [\theta' + 2(p-1)\alpha], \tag{90}$$

где n' — показатель преломления пленки, а n — показатель преломления окружающей среды. Виртуальные волновые фронты прошедшей группы W_1, W_2, ..., W_p, ..., в плоскости которых лежат ребра клина, были бы синфазны, если бы при отражении не происходило изменения фазы. В точке P на второй поверхности на расстоянии ρ от O оптическая разность хода между p-й волной и волной, прошедшей прямо ($p = 1$), равна

$$\Delta\mathcal{S}_p = n(PN_p - PN_1) = n\rho(\sin\theta_p - \sin\theta_1), \tag{91}$$

где N_p и N_1 — основания перпендикуляров, опущенных из точки P на W_p и W_1 соответственно. Если φ — изменение фазы при однократном отражении на любой поверхности клина, то для полной разности фаз δ_p между p-й волной

*) Приводимый ниже анализ принадлежит Бросселу [84].

и волной, прошедшей прямо, получим, учитывая (90) и (91),

$$\delta_p=\frac{2\pi}{\lambda_0}\Delta\mathcal{S}_p+2(p-1)\varphi=\frac{2\pi}{\lambda_0}n'\rho\{\sin[\theta'+2(p-1)\alpha]-\sin\theta'\}+2(p-1)\varphi=$$
$$=\frac{4\pi}{\lambda_0}n'h\cos\theta'\frac{\sin(p-1)\alpha}{\operatorname{tg}\alpha}\{\cos(p-1)\alpha-\operatorname{tg}\theta'\sin(p-1)\alpha\}+2(p-1)\varphi, \qquad (92)$$

где

$$h=\rho\operatorname{tg}\alpha$$

— толщина пленки в точке P. Когда прошедшая волна падает на вторую поверхность со стороны нормали, удаленной от ребра клина, волновой фронт W_1 действительный; в этом случае δ_p получается из (92), но знак члена с $\operatorname{tg}\theta'$ изменяется.

В обозначениях, принятых в п. 7.6.1, амплитуды прошедших волн имеют вид

$$tt'A^{(i)}\exp(i\delta_1),\quad tt'r'^2A^{(i)}\exp(i\delta_2),\quad \ldots,\quad tt'r'^{2(p-1)}A^{(i)}\exp(i\delta_p),\ \ldots$$

Амплитуда в P, получающаяся в результате суперпозиции бесконечного ряда таких волн, равна

$$A^{(t)}=A^{(i)}tt'\sum_{p=1}^{\infty}r'^{2(p-1)}\exp(i\delta_p)=A^{(i)}\mathcal{T}\sum_{p=1}^{\infty}\mathcal{R}^{p-1}\exp(i\delta_p), \qquad (93)$$

а соответствующая интенсивность равна

$$I^{(t)}=I^{(i)}\mathcal{T}^2\left|\sum_{p=1}^{\infty}\mathcal{R}^{p-1}\exp(i\delta_p)\right|^2. \qquad (94)$$

В дальнейшем исследовании сохраним только конечное число членов в рядах (93) и (94). $|A^{(t)}|$ не может превышать $|A^{(i)}|\mathcal{T}/(1-\mathcal{R})$ и, если учитываются только первые p членов, ошибка $|\Delta A^{(t)}|$ в $A^{(t)}$ не может превысить $|A^{(i)}|\mathcal{T}\mathcal{R}^{(p)}/(1-\mathcal{R})$. Соответствующая относительная ошибка $|\Delta I^{(t)}|/I^{(t)}$ удовлетворяет соотношению

$$\frac{|\Delta I^{(t)}|}{I^{(t)}}\leqslant 2\left|\frac{\Delta A^{(t)}}{A^{(t)}}\right|\leqslant 2\left|\frac{A^{(i)}}{A^{(t)}}\right|\frac{\mathcal{T}\mathcal{R}^p}{(1-\mathcal{R})}. \qquad (95)$$

Далее мы увидим, что в условиях, представляющих наибольший практический интерес, $I^{(t)}$ достаточно точно определяется формулой Эйри (13), согласно которой минимум интенсивности соответствует $|A^{(t)}|=A^{(i)}\mathcal{T}/(1+\mathcal{R})$. Следовательно, если мы хотим, чтобы для этого частного значения $|A^{(t)}|$ величина $|\Delta I^{(t)}|/I^{(t)}$ не превышала, например, 0,01, что дает относительную ошибку в определении интенсивности по соседству с максимумом $I^{(t)}$, значительно меньшую чем 0,01, то, согласно (95), число p членов, которые нужно оставить в ряде, получается из соотношения

$$\mathcal{R}^p\approx\frac{1}{200}\frac{1-\mathcal{R}}{1+\mathcal{R}}. \qquad (96)$$

Например, если $\mathcal{R}=0{,}93$, то $p\approx 120$.

Только при конечном числе членов в ряду (94) и при достаточно малом α (92) можно разложить по степеням α. Сохраняя в этом разложении члены со степенью не выше второй, получим

$$\delta_p=(p-1)\left\{\frac{4\pi}{\lambda_0}n'h\cos\theta'+2\varphi\right\}-(p-1)^2\alpha\frac{4\pi}{\lambda_0}n'h\sin\theta'-$$
$$-\frac{(p-1)(2p^2-4p+3)\alpha^2}{3}\frac{4\pi n'h\cos\theta'}{\lambda_0}. \qquad (97)$$

В частности, при нормальном падении ($\theta'=0$) имеем

$$\delta_p=(p-1)\left(\frac{4\pi n'h}{\lambda_0}+2\varphi\right)-\frac{(p-1)(2p^2-4p+3)\alpha^2}{3}\cdot\frac{4\pi n'h}{\lambda_0}. \qquad (98)$$

Можно пренебречь членом с α^2, если он мал по сравнению с π, и считать разность фаз $\delta_p - \delta_{p-1}$, следующих друг за другом волн постоянной и равной $\delta = 4\pi n'h/\lambda_0 + 2\varphi$. В таком случае $I^{(t)}$ определяется формулой Эйри (13), и в плоскости клина наблюдаются полосы с таким же распределением интенсивности, как и у локализованных в бесконечности полос, создаваемых плоско-параллельной пластинкой (см. рис. 7.58). Положения максимумов интенсивности, соответствующие $\delta = 2m\pi$, определяются из соотношения

$$2n'h = \left(m - \frac{\varphi}{\pi}\right)\lambda_0, \qquad m = 1,\ 2,\ \ldots \qquad (99)$$

Следовательно, эти полосы параллельны ребру клина и находятся на расстоянии $\lambda_0/2n'\alpha$ друг от друга. Если отражательная способность велика ($\mathcal{R} \gtrsim 0{,}9$) и, значит, велико p, то мы вправе заменить $(p-1)(2p^2-4p+3)$ на $2p^3$ и условие малости члена с α^2 в (98) по сравнению с π дает

$$n'h \ll \frac{3\lambda_0}{8p^3\alpha^2}. \qquad (100)$$

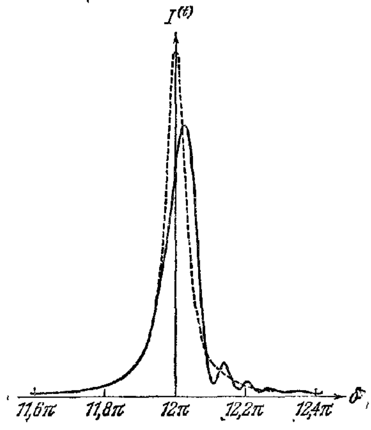


Рис 7.74. Сравнение распределения интенсивности в многолучевых интерференционных полосах Физо (сплошная кривая) с соответствующим распределением интенсивности Эйри (пунктир) [85].

При величинах α, соответствующих расстоянию между соседними полосами порядка 1 *см* или меньше, это условие строго ограничивает $n'h$, и распределение интенсивности полос совпадает с распределением Эйри только вблизи ребра клина, где оптическая толщина составляет несколько длин волн. Например, при $\alpha \approx 2{,}5 \cdot 10^{-4}$ (что при $n' \sim 1$ и $\lambda_0 = 5500$ Å соответствует примерно одной полосе на 1 *мм*) $p \sim 50$ ($\mathcal{R} \sim 0{,}9$) и условие (100) дает $n'h \ll 50\lambda_0$. При большем удалении от ребра клина член с α^2 в (98) значительно возрастает и численный расчет показывает (рис. 7.74), что на значение $I^{(t)}$ это влияет следующим образом: интенсивность максимумов становится меньше, а их полуширина больше, чем следует из формулы Эйри; максимумы смещаются от положений, определяемых (99), в сторону от ребра клина; полосы становятся асимметричными за счет появления вторичных максимумов на стороне, удаленной от ребра клина.

Согласно (99) интерференционные полосы соответствуют линиям равной оптической толщины в случае плоских поверхностей пленки. Однако важное практическое применение интерференционных полос заключается в исследовании пленок с неплоскими поверхностями; поэтому сейчас мы должны выяснить, насколько строго в этих условиях формы полос соответствуют контурам равной толщины. У полос, точно отражающих нарушения постоянства толщины пленки, интенсивность в каждой точке P должна однозначно зависеть от толщины пленки h в P, что обычно не имеет места. Луч, приходящий в точку P после $2(p-1)$ отражений (рис. 7.75, *а*) входит в пленку в A_1 и отражается в точках $B_1, C_1, \ldots, B_{p-1}, C_{p-1}$. Его вклад в результирующую амплитуду в точке P зависит от оптической толщины пленки между A_1 и P. Таким образом, полосы могут быть смещены относительно положения, соответствующего истинным очертаниям контуров равной оптической толщины, на величину, зависящую от формы пленки близ P. Этот эффект, конечно, наиболее существен при высокой отражательной способности, но при данном отражении он минимален там, где толщина пленки меньше всего и падение света близко к нормальному. Можно оценить размер участка пленки, который оказывает влия-

ние на интенсивность в P, если предположить, что он имеет плоскую поверхность (рис. 7.75, *б*). Пусть $A'_2, A''_2, \ldots, A'_{p-1}, A''_{p-1}$ — изображения A_1, соответствующие ряду последовательных отражений на второй и первой поверхностях пленки. Эти изображения, находящиеся в главном сечении клина, проходящем через A_1, лежат на окружности с центром в O на ребре клина и радиусом $OA_1 = \rho'$. Угол $A_1OA''_p$ равен $2(p-1)\alpha$. Изображение A''_{p-1} связано с отрезком луча $C_{p-1}P$ и поэтому лежит на его продолжении. Допустим, что луч, приходящий в B_2, лежит в главном сечении клина и падает под углом θ' с той стороны

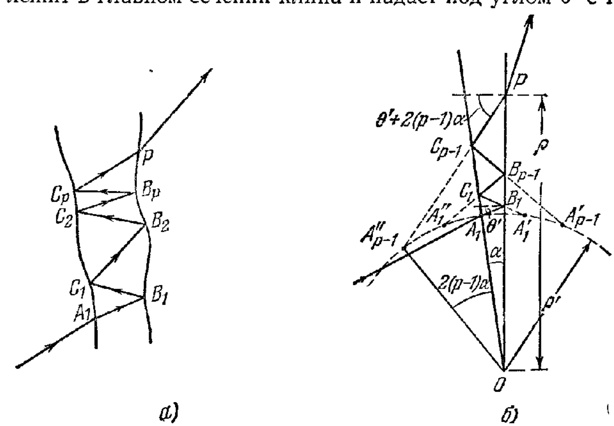


Рис. 7.75. Смещение многократно отраженного луча по клинообразной пленке.
а — поверхности неправильной формы; *б* — плоские поверхности.

нормали, где находится точка O; тогда угол падения этого луча, когда он достигает P, равен $\theta' + 2(p-1)\alpha$. В треугольнике $A''_{p-1}PO$ сторона $OA''_{p-1} = \rho'$, и мы можем написать

$$\angle A''_{p-1}PO = \frac{\pi}{2} - [\theta' + 2(p-1)\alpha], \qquad \angle PA''_{p-1}O = \frac{\pi}{2} + (\theta' - \alpha);$$

следовательно, полагая $OP = \rho$, получим $\rho'/\cos[\theta' + 2(p-1)\alpha] = \rho/\cos(\theta' - \alpha)$. Поэтому

$$\Delta\rho = \rho - \rho' = \frac{h}{\operatorname{tg}\alpha}\left\{1 - \frac{\cos[\theta' + 2(p-1)\alpha]}{\cos(\theta' - \alpha)}\right\}. \qquad (101)$$

Полагая $\theta' = 0$ и пренебрегая членами со степенью α, большей второй, получим

$$\Delta\rho = \frac{4p^2 - 8p + 3}{2} h\alpha. \qquad (102)$$

Следовательно, если существует p достаточно интенсивных пучков, то можно ожидать, что влияние толщины пленки будет сказываться на интенсивности в P на расстояниях порядка $2p^2h\alpha$ от P. Это расстояние мало даже при высокой отражательной способности поверхностей, если величина h достигает лишь нескольких длин волн видимого света. Например, при $p \sim 50$, $h = 2{,}5 \cdot 10^{-4}$ *см* (около 5 длин волн) и $\alpha = 2{,}5 \cdot 10^{-4}$ оно составляет около $3 \cdot 10^{-3}$ *мм*. Если внимательно интерпретировать малые искажения полос, то можно считать, что полосы соответствуют контурам равных оптических толщин неоднородных пленок в тех случаях, когда толщина пленки достаточно мала.

До сих пор предполагалось, что источник света точечный, но практически из-за недостатка света следует использовать источники максимально допустимых размеров. Отбрасывая в (97) члены с α и α^2, получим для положения

максимумов интенсивности в свете, падающем не под прямым углом, соотношение

$$2n'h\cos\theta' = \left(m - \frac{\varphi}{\pi}\right)\lambda_0, \qquad m = 1, 2, \ldots \tag{103}$$

Здесь мы пренебрегаем углом клина, и поэтому оно остается справедливым для любой плоскости падения. Следовательно, при работе с источником, протяженным в радиальном направлении, т. е. при углах падения, находящихся в пределах $0 \leqslant \theta' \leqslant \varepsilon'$, максимумы данного порядка занимают участок, соответствующий Δm порядкам, и, согласно (103), для малого ε' находим

$$\Delta m = \frac{2}{\lambda_0} n'h \,|\Delta(\cos\theta')| = \frac{n'h\varepsilon'^2}{\lambda_0}. \tag{104}$$

Такое расширение, связанное с увеличением размера источника, зависит от резкости $\mathcal{F}$ полос, образованных единичным элементом источника; если $n'h\varepsilon'^2/\lambda_0 \ll 1/\mathcal{F}$, то увеличение ε' влияет главным образом на увеличение интенсивности максимумов; если же $n'h\varepsilon'^2/\lambda_0 \gg 1/\mathcal{F}$, то влияние ε' сказывается главным образом на ширине полос. Данный эффект формально очень напоминает расширение полос в интерферометре Фабри — Перо, связанное с наличием у пластин сферической кривизны (см. стр. 305), и это может помочь нам в его изучении. Соответствующая обработка результатов исследований Дюфура и Пикка [54] показывает, что если $n'h\varepsilon'^2/\lambda_0$ превышает $1/\mathcal{F}$, то наблюдается лишь небольшое увеличение интенсивности максимумов, и можно утверждать, что при достижении половины этой величины полосы будут примерно на 10% шире, чем при точечном источнике. Учитывая преломление лучей на первой поверхности пленки, можно определить допустимый угловой радиус ε источника соотношением

$$\varepsilon \approx \frac{n'\varepsilon'}{n} \approx \frac{1}{n}\sqrt{\frac{\lambda_0 n'}{2h\mathcal{F}}}. \tag{105}$$

Его можно сравнить с аналогичным соотношением (7.5.73), относящимся к интерференции двух пучков. Если величина h составляет только несколько длин волн, ε не очень мало; например, при $h \approx 5\lambda_0$, $\mathcal{F} = 30$ ($\mathcal{R} \approx 0{,}9$), $n' \approx n \approx 1$, из (105) получим $\varepsilon \approx 3^\circ$. Следовательно, падающий свет не должен быть строго параллельным, и для наблюдения полос с большим увеличением протяженный источник имеет преимущество.

Если свет квазимонохроматичен и длины волн его компонент распределены в области $\Delta\lambda_0$ около средней длины волны $\bar{\lambda}_0$, то максимумы данного порядка занимают участок, соответствующий Δm порядкам. Пренебрегая зависимостью n' и φ от длины волны, получим из (99) $\Delta m = 2n'h\Delta\lambda_0/\bar{\lambda}_0^2$. Полосы незначительно расширяются при отступлении от строгой монохроматичности, если величина Δm мала по сравнению с $1/\mathcal{F}$, т. е. если

$$\Delta\lambda_0 \ll \frac{(\bar{\lambda}_0)^2}{2n'h\mathcal{F}}. \tag{106}$$

Если $n'h$ составляет несколько длин волн, последнее условие нельзя считать жестким; например, при $n'h \approx 5\lambda_0$, $\mathcal{F} = 30$, $\lambda_0 = 5500$ Å (106) дает $\Delta\lambda_0 \ll 18$ Å. Следовательно, источник не должен принадлежать к типу, применяемому в интерферометрии при больших разностях, и можно пользоваться обычным дуговым разрядом высокого давления, создающим большую яркость и позволяющим производить наблюдение при большом увеличении.

До сих пор мы имели дело со светом, проходящим сквозь пленку, однако многолучевую интерференционную картину можно наблюдать и в отраженном свете. Она оказывается дополнительной к картине в проходящем свете, если на отражающих поверхностях не происходит поглощения. Такую дополнительность надо понимать в том смысле, что для обеих картин сумма интенсивностей в каждой точке равна интенсивности падающего света. Влияние поглощения,

которое практически всегда существует у отражающих покрытий, рассмотрено в работе [86]. В ней показано, что при использовании хорошо отражающих и мало поглощающих покрытий из серебра интерференционная картина в отраженном свете состоит из темных полос на почти равномерном светлом фоне, причем эти полосы уже соответствующих светлых полос в проходящем свете, но в отраженном свете интенсивность в минимумах очень чувствительна к поглощающим свойствам покрытия на первой поверхности пленки, и если поглощение слишком велико, то полосы исчезают.

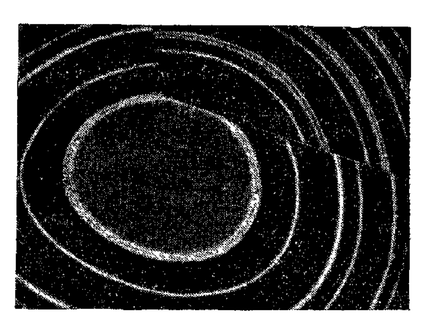


Рис. 7.76. Многолучевые интерференционные полосы Физо на поверхности скола слюды в свете зеленой (5461 Å) и желтых (5770 Å и 5790 Å) линий ртути (проходящий свет). (По Вилкоку).

Многолучевыми полосами Физо пользуются в оптических цехах для испытания высококачественных оптических поверхностей, например у пластин, применяемых в интерферометре Фабри — Перо; они широко использовались также Толанским и его сотрудниками при изучении топографии почти плоских кристаллических и металлических поверхностей [87]. Испытуемая и оптически плоская эталонная поверхности, покрытые отражающими слоями серебра, плотно прижимаются друг к другу. Создаваемые воздушной прослойкой интерференционные полосы рассматриваются в микроскоп с большой угловой апертурой, позволяющей использовать все полезные пучки. Если $\lambda = \lambda_0/n'$ — длина волны в воздухе, то полосы соответствуют контурам исследуемой поверхности, определяемым плоскостями, параллельными эталонной поверхности и разделенными интервалами $\lambda/2$. При достаточно большом клине между пластинами в поле зрения появляется большое число полос (см., например, рис. 7.76), и можно измерить неровности исследуемой поверхности, определяя отклонение полос от прямых линий; там, где неровности поверхности достигают Δh, полосы смещаются в сторону на Δm порядков; так как

$$\Delta h = \frac{\lambda}{2}\Delta m, \tag{107}$$

то можно определить Δh, если известно Δm. Величина Δm приблизительно равна отношению смещения данной полосы к расстоянию между соседними полосами. При большой отражательной способности покрытий таким способом можно измерять очень малые неровности исследуемой поверхности. Легко измеряемое смещение на полуширину полосы соответствует $\Delta h = \lambda/2\mathcal{F}$; при $\lambda = 5500$ Å и $\mathcal{F} = 40$ это составляет всего лишь 70 Å *). Наблюдая за направлением смещения полос при изменении расстояния между пластинами, часто можно решить вопрос о том, имеем ли мы на поверхности углубление или выпуклость. Однако, вообще говоря, для этой цели необходимы наблюдения в нескольких длинах волн. При наблюдении в проходящем свете можно использовать одновременно несколько длин волн; особенно удобны желтые линии ртути с длинами волн $\lambda = 5770$ Å и $\lambda = 5790$ Å. В этом свете полосы расположены тесными парами, расстояние между которыми увеличивается по мере увеличения порядка интерференции; поэтому легко определяется тип неровности поверхности. Если же наблюдаются разрывы полос, например, при на-

*) Толщина серебряного покрытия достигает 500 Å, однако установлено, что оно не изменяет топографии поверхности, по крайней мере настолько, чтобы это удалось обнаружить подобным методом.

личии ступеньки (плоскости скола) в кристалле, то по изменению порядка можно также определить характер поверхности кристалла.

Если исследуемая поверхность близка к плоскости, то, уменьшая ее наклон относительно эталонной оптически плоской поверхности, можно добиться того, чтобы все поле зрения было занято только одной полосой (см., например, рис. 7.77). В этих условиях крайне малые изменения высоты приводят к заметным изменениям интенсивности. Пусть I и $I+\Delta I$ — интенсивности участков, обусловленные соседними площадками, отличающимися по высоте на малую величину Δh, которым соответствует изменение δ на $\Delta\delta$. Контрастность, определяемая как $|\Delta I|/I$, максимальна, когда δ выбрано таким, чтобы $|\partial I/\partial\delta|/I$ было максимальным. Рассматривая картину в проходящем свете, найдем из (27)



Рис. 7.77. Многолучевая интерференционная полоса Физо на поверхности кристалла алмаза. (По Вилкоку).

$$\frac{I^{(t)}}{I^{(t)}_{\text{макс}}}=\frac{1}{1+F\sin^2\frac{\delta}{2}}. \tag{108}$$

Отсюда

$$\frac{\partial}{\partial\delta}\left(\frac{I^{(t)}}{I^{(t)}_{\text{макс}}}\right)=\frac{-F\sin\delta}{2\left(1+F\sin^2\frac{\delta}{2}\right)^2}. \tag{109}$$

Следовательно,

$$\frac{1}{I^{(t)}}\left|\frac{\partial I^{(t)}}{\partial\delta}\right|=\frac{F\sin\delta}{2\left(1+F\sin^2\frac{\delta}{2}\right)}, \tag{110}$$

так что при максимальной контрастности имеем

$$2\left(1+F\sin^2\frac{\delta}{2}\right)\cos\delta-F\sin^2\delta=0. \tag{111}$$

При высокой отражательной способности F велико, и значения δ, удовлетворяющие (111), близки к целым кратным 2π; поэтому можно написать $\delta=2m\pi\pm\varepsilon$, где m — целое, а ε — малая величина по сравнению с $\pi/2$. В таком случае из (111), пренебрегая степенями ε, большими второй, получим

$$\varepsilon=\frac{2}{\sqrt{F+2}}\approx\frac{2}{\sqrt{F}}, \tag{112}$$

и (110) сводится к

$$\frac{1}{I^{(t)}}\left|\frac{\partial I^{(t)}}{\partial\delta}\right|_{\text{макс}}\approx\frac{F\varepsilon}{2\left(1+\frac{F\varepsilon^2}{4}\right)}\approx\frac{\sqrt{F}}{2}. \tag{113}$$

Следовательно, в первом приближении максимум контрастности определяется выражением

$$\left|\frac{\Delta I^{(t)}}{I^{(t)}}\right|_{\text{макс}}=\frac{1}{I^{(t)}}\left|\frac{\partial I^{(t)}}{\partial\delta}\right|_{\text{макс}}|\Delta\delta|\approx\frac{\sqrt{F}}{2}|\Delta\delta|. \tag{114}$$

Так как в соответствии с (99) $\Delta\delta$ и Δh связаны соотношением $|\Delta\delta|=4\pi|\Delta h|\lambda$, то, используя (22), можно написать для максимума контрастности

$$\left|\frac{\Delta I^{(t)}}{I^{(t)}}\right|_{\text{макс}}\approx\frac{4\mathcal{F}}{\lambda}|\Delta h|. \tag{115}$$

При $\lambda=5500$Å и $\mathcal{F}=40$ последнее соотношение дает $|\Delta I^{(t)}/I^{(t)}|_{\text{макс}}\approx 0{,}1$, если $\Delta h\approx 3{,}5$Å, что порядка молекулярных размеров.

До сих пор мы рассматривали эффекты в квазимонохроматическом свете. Теперь предположим, что пленка освещается белым светом, падающим нор-

мально, и допустим, что в проходящем свете ахроматическая линза L дает изображение пленки на плоскости щели спектрографа (рис. 7.78). Для каждой спектральной компоненты фазовые соотношения пучков, приходящих в точку P' на щели спектрографа, такие же, как и в точке пленки P, сопряженной с P'.

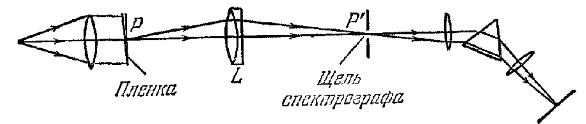


Рис. 7.78. Схема установки для наблюдения интерференционных полос, полученных с тонкой пленкой, в белом свете.

Следовательно, если плёнка достаточно тонка и удовлетворяет условию (100), то, согласно (99), существуют максимумы интенсивности в точке P' для тех длин волн $\lambda_0^{(m)}$, которые удовлетворяют соотношению

$$\lambda_0^{(m)}=\frac{2n'h}{m-\varphi/\pi},\qquad m=1,\ 2,\ \ldots,\tag{116}$$

где $n'h$— оптическая толщина пленки в точке P. Интенсивность для длин волн между этими величинами получается из формулы Эйри. Таким образом, при высокой отражательной способности пленки спектр в фокальной плоскости спектрографа пересечен узкими яркими полосами, разделяющимися значительно более широкими темными промежутками, и различие в длинах волн соседних полос тем больше, чем тоньше пленка.

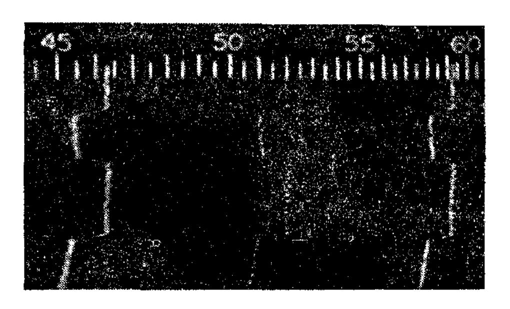


Рис. 7.79. Полосы равного хроматического порядка, образованные на участке поверхности кристалла алмаза [89].

Шкала показывает длины волн в сотнях ангстрем.

В особом случае, когда величина $n'h$ постоянна, $\lambda_0^{(m)}$ также постоянна, и полосы представляют собой прямые линии, параллельные щели. Их иногда называют *полосами Эдсера — Бутлера*. Если φ не зависит от длины волны, полосы располагаются через равные интервалы $\Delta\kappa_0=1/2n'h$ ($\kappa_0=1/\lambda_0$). Эта особенность делает их полезными для калибровки спектрографов, особенно в инфракрасной области, где эталонные длины волн немногочисленны. Обычно $\lambda_0^{(m)}$ меняется вдоль полосы в зависимости от изменения оптической толщины вдоль сечения пленки, сопряженного со щелью. В частности, если пленка представляет собой воздушный зазор между исследуемой и оптически плоской поверхностями, такие полосы позволяют с большой точностью найти профиль выбранного сечения исследуемой поверхности. Эти полосы, впервые описанные Толанским [88], называют иногда *полосами равного хроматического* порядка (рис. 7.79).

Пусть толщины пленки в двух точках P_1 и P_2 равны h_1 и h_2, и пусть в сопряженных точках P_1' и P_2' на щели имеются максимумы порядка m в длинах волн в воздухе $\lambda_1^{(m)}$ и $\lambda_2^{(m)}$ соответственно. Так как $\lambda^{(m)}=\lambda_0^{(m)}/n'$, то из (116) следует

$$\Delta h=h_2-h_1=\left(m-\frac{\varphi_1}{\pi}\right)\frac{\lambda_2^{(m)}-\lambda_1^{(m)}}{2}+\frac{a}{2}\lambda_2^{(m)},\tag{117а}$$

где

$$a=\frac{\varphi_1-\varphi_2}{\pi}, \qquad (117б)$$

а φ_1 и φ_2 — сдвиги фаз при отражении для длин волн $\lambda_1^{(m)}$ и $\lambda_2^{(m)}$. Величину $m-\varphi_1/\pi$ можно найти, измеряя расстояние в длинах волн между соседними полосами. Так, если для точки P_1' на щели имеется максимум порядка $m+1$ в длине волны $\lambda_1^{(m+1)}$, то, учитывая (116), получим

$$\left(m-\frac{\varphi_1}{\pi}\right)\lambda_1^{(m)}=\left(m+1-\frac{\varphi_1'}{\pi}\right)\lambda_1^{(m+1)},$$

т. е.

$$m-\frac{\varphi_1}{\pi}=(1+b)\frac{\lambda_1^{(m+1)}}{\lambda_1^{(m)}-\lambda_1^{(m+1)}}, \qquad (118а)$$

где

$$b=\frac{\varphi_1-\varphi_1'}{\pi}, \qquad (118б)$$

а φ_1' — изменение фазы при отражении волны длиной $\lambda_1^{(m+1)}$. В том случае, когда полосы прерываются, для идентификации соответствующих порядков нужно измерить расстояние в длинах волн между соседними полосами с каждой стороны от разрыва. Из (117) и (118) имеем

$$\Delta h=(1+b)\frac{\lambda_1^{(m+1)}}{\lambda_1^{(m)}-\lambda_1^{(m+1)}}\,\frac{\lambda_2^{(m)}-\lambda_1^{(m)}}{2}+\frac{a}{2}\lambda_2^{(m)}. \qquad (119)$$

Можно пренебречь величинами a и b, появляющимися вследствие вариации сдвига фазы с длиной волны при отражении на серебряных покрытиях, если рассматриваемый интервал длин волн не слишком велик. В таком случае измерения $\lambda_1^{(m)}$, $\lambda_1^{(m+1)}$ и $\lambda_2^{(m)}$ дают Δh, и так как по (119) Δh пропорционально $\lambda_2^{(m)}-\lambda_1^{(m)}$, то профиль выбранного сечения исследуемой поверхности определяет одна полоса.

7.6.8. Многолучевые полосы, получающиеся с двумя плоскопараллельными пластинками. *а. Полосы в монохроматическом и квазимонохроматическом свете.* Две плоскопараллельные пластинки с хорошо отражающими поверхностями, расположенные друг за другом. На пластинки падают плоские волны монохроматического света (рис. 7.80). Воспользуемся обозначениями, принятыми в п. 7.6.1, и отметим пластинки индексами 1 и 2. Интенсивность света, прошедшего сквозь первую пластинку, согласно уравнениям (26) и (27), запишется в виде

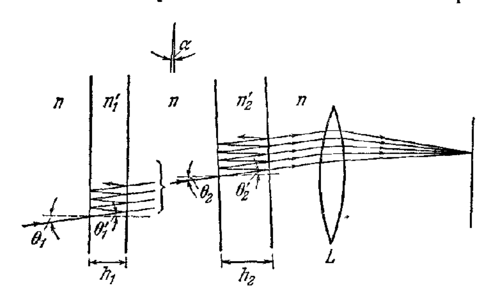


Рис. 7.80. К возникновению интерференционных полос, локализованных в бесконечности, после многократных отражений в двух плоскопараллельных пластинках.

$$I_1^{(t)}=\frac{\mathcal{T}_1}{(1-\mathcal{R}_1)^2}\,\frac{1}{1+F_1\sin^2\frac{\delta_1}{2}}I^{(i)}, \qquad (120)$$

где $I^{(i)}$ — интенсивность падающего света. Аналогично интенсивность света, прошедшего сквозь вторую пластинку, соответствующая интенсивности падающего света $I_1^{(t)}$, выразится соотношением

$$I_2^{(t)}=\frac{\mathcal{T}_2^2}{(1-\mathcal{R}_2)^2}\,\frac{1}{1+F_2\sin^2\frac{\delta_2}{2}}I_1^{(t)}. \qquad (121)$$

Предположим, что свет, отразившийся от второй пластинки обратно и снова отразившийся вперед от первой, исключается. Это означает, что расстояние между пластинками достаточно велико по сравнению с их рабочей поверхностью, и если пластинки взаимно параллельны, то направление падения света не слишком близко к нормальному. Полная интенсивность $I^{(t)}$ прошедшего света тогда равна $I_2^{(t)}$, и согласно (120) и (121) имеем

$$\frac{I^{(t)}}{I^{(i)}}=\frac{\mathcal{T}_1^2}{(1-\mathcal{R}_1)^2}\frac{\mathcal{T}_2^2}{(1-\mathcal{R}_2)^2}\frac{1}{\left(1+F_1\sin^2\frac{\delta_1}{2}\right)}\frac{1}{\left(1+F_2\sin^2\frac{\delta_2}{2}\right)}. \qquad (122)$$

Следовательно, в фокальной плоскости линзы L образуются полосы с относительным распределением интенсивности, равным произведению относительных распределений интенсивностей в полосах, создаваемых каждой пластинкой.

Разности фаз δ_1 и δ_2, согласно (25), равны

$$\delta_1=\frac{4\pi}{\lambda_0}n_1'h_1\cos\theta_1'+2\varphi_1, \qquad \delta_2=\frac{4\pi}{\lambda_0}n_2'h_2\cos\theta_2'+2\varphi_2. \qquad (123)$$

Рассмотрим случай, когда две пластинки взаимно параллельны, так что системы колец, которые получались бы с каждой пластинкой отдельно, концентричны. Когда угол падения $\theta_1=\theta_2=\theta$ мал, $\cos\theta_1'\approx 1-\theta_1'^2/2$, $\cos\theta_2'\approx 1-\theta_2'^2/2$ и если n — показатель преломления окружающей среды, то по закону преломления получим $\theta_1'\approx n\theta/n_1'$, $\theta_2'\approx n\theta/n_2'$. Следовательно, из (123) с точностью до второго порядка по θ имеем

$$\delta_1=\frac{4\pi}{\lambda_0}h_1\left(n_1'-\frac{n^2\theta^2}{2n_1'}\right)+2\varphi_1, \qquad \delta_2=\frac{4\pi}{\lambda_0}h_2\left(n_2'-\frac{n^2\theta^2}{2n_2'}\right)+2\varphi_2. \qquad (124)$$

Отсюда вытекает, что при изменении θ соответствующие изменения δ_1 и δ_2 находятся в отношении

$$\frac{\Delta\delta_2}{\Delta\delta_1}=\frac{n_1'h_2}{n_2'h_1}, \qquad (125)$$

причем φ считается не зависящим от θ. Если показатели преломления и толщины пластинок таковы, что в некотором направлении $\theta=\theta_0$, порядок интерференции для каждой пластины равен целому числу, т. е. если

$$\theta=\theta_0, \quad \delta_1=2m_1\pi, \quad \delta_2=2m_2\pi \quad (m_1 \text{ и } m_2 \text{— целые}), \qquad (126)$$

то оба члена с F_1 и F_2 в (122) достигают максимального значения, равного единице, и при $\theta=\theta_0$ существует абсолютный максимум интенсивности прошедшего света. Если, кроме того,

$$\frac{n_1'h_2}{n_2'h_1}=a, \qquad (127)$$

a — положительное целое число, то из (125) видно, что целые порядки (m_2-a), (m_2-2a), ... для второй пластинки соответствуют тем же величинам θ, что и целые порядки (m_1-1), (m_1-2), ... для первой пластинки. Следовательно, и в этих направлениях также существуют абсолютные максимумы $I^{(t)}/I^{(i)}$.

Практически требуется только ограниченное число p совпадений, и тогда $n_1'h_2/n_2'h_1$ достаточно близко к целому a, если взаимное смещение максимумов, обусловленных членами с F_1 и F_2 при $\delta_1=2(m_1-p)\pi$ и $\delta_2=2(m_2-pa)\pi$, мало по сравнению с их полушириной. Используя (125) и уравнение (21), найдем для этих полуширин, выраженных через δ_2, величины $4n_1'h_2/n_2'h_1\sqrt{F_1}$ и $4/\sqrt{F_2}$. Следовательно, необходимое условие имеет вид

$$\left|\frac{n_1'h_2}{n_2'h_1}2p\pi-2pa\pi\right|\ll\frac{n_1'h_2}{n_2'h_1}\frac{4}{\sqrt{F_1}}+\frac{4}{\sqrt{F_2}},$$

которое можно переписать для $(n_1'h_2/n_2'h_1)\approx a$ в виде

$$\left|\frac{n_1'h_2}{n_2'h_1}-a\right|\ll\frac{1}{p}\left(\frac{a}{\mathcal{F}_1}+\frac{1}{\mathcal{F}_2}\right), \qquad (126a)$$

где $\mathcal{F}_1$ и $\mathcal{F}_2$ — резкости полос, получающихся с каждой пластиной в отдельности. Далее, до тех пор, пока p не слишком велико и порядок интерференции m_1 для первой пластины не слишком мал, соотношения (126) и (126а) могут практически удовлетворяться одновременно. Поэтому при $\mathcal{F}_1=\mathcal{F}_2=30$ и $a=1$ (126а) требует, чтобы $|(n_1'h_2/n_2'h_1)-a|\ll 1/15\ p$, тогда как (126) всегда может удовлетворяться при изменении $n_2'h_2$ на $\lambda_0/4$ и соответствующее изменение $n_1'h_2/n_2'h_1$ составит около $n_1'\lambda_0/4n_2'^2h_1\approx n_1^2/2n_2'^2m_1$, так как $m_1\approx 2\,n_1'h_1/\lambda_0$.

Вид кривой интенсивности прошедшего света показан на рис. 7.81. Между последовательными главными максимумами находятся $(a-1)$ вторичных максимумов, соответствующих целым порядкам (m_2-1), (m_2-2), ..., (m_2-a+1), ..., от второй пластинки. Если F_1 велико, а a не слишком велико, то соответствующие величины членов с F_1 малы по сравнению с единицей, и вторичные максимумы много слабее главных.

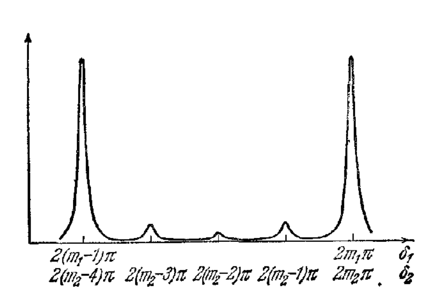


Рис. 7.81. Отношение интенсивности прошедшего света к интенсивности падающего как функция разностей фаз δ_1 и δ_2.

Многолучевые интерференционные полосы равного наклона в двух плоскопараллельных пластинках. $\mathcal{R}_1=0{,}64$, $\mathcal{R}_2=0{,}64$, $a=4$.

Полосы такого рода, полученные с двумя интерферометрами Фабри — Перо, используются для исследования тонкой структуры спектральных линий. Разделяющие кольца выбирают так, чтобы h_2/h_1 приблизительно равнялось подходящему целому a, а точная установка осуществляется изменением давления воздуха (и, следовательно, его показателя преломления) в герметичном контейнере, в который помещен один из интерферометров *). При a, большем единицы, такая установка имеет преимущество перед одиночным интерферометром, так как при данной разрешающей силе обладает большей областью дисперсии **). Изменение длины волны, необходимое для того, чтобы сместить картину на величину, равную расстоянию между соседними главными максимумами, соответствует изменению на единицу порядка интерференции в первой пластинке и изменению на a порядка интерференции во второй пластине. Следовательно, область дисперсии, в которой не происходит перекрывания главных максимумов, равняется области дисперсии интерферометра с меньшим расстоянием между пластинами и в a раз больше, чем у интерферометра с бо́льшим расстоянием между пластинами (см. (46)). Полуширина главных максимумов соответствует $\delta_1=2m_1\pi\pm\varepsilon_1/2$ или $\delta_2=2m_2\pi\pm\pm\varepsilon_2/2$ (m_1, m_2 — целые числа), где, согласно (122),

$$\left(1+F_1\sin^2\frac{\varepsilon_1}{4}\right)\left(1+F_2\sin^2\frac{\varepsilon_2}{4}\right)=2,$$

или в соответствии с (125) и (127)

$$\left(1+F_1\sin^2\frac{\varepsilon_2}{4a}\right)\left(1+F_2\sin^2\frac{\varepsilon_2}{4}\right)=2. \tag{128}$$

При достаточно большом F_2 величина ε_2 значительно меньше $\pi/2$ и можно принять $\sin\varepsilon_2=\varepsilon_2$. Тогда (128) сведется к

$$\left(\frac{4}{\varepsilon_2}\right)^4-\left(F_2+\frac{F_1}{a^2}\right)\left(\frac{4}{\varepsilon_2}\right)^2-\frac{F_1F_2}{a^2}=0. \tag{129}$$

*) Точную установку (хотя и менее совершенную) можно также осуществить, наклоняя один из интерферометров относительно другого.

**) Такое устройство было впервые описано в [90].

откуда

$$\varepsilon_2 = \frac{4\sqrt{2}}{\sqrt{(F_2 + F_1/a^2) + \sqrt{F_2^2 + 6F_1F_2/a^2 + F_1^2/a^4}}}. \tag{130}$$

Если удалить интерферометр с меньшим расстоянием между пластинами, то $F_1 = 0$ и последнее соотношение сведется к $\varepsilon_2 = 4/\sqrt{F_2}$, что находится в согласии с (21). В других случаях ε_2 меньше $4/\sqrt{F_2}$, поэтому полуширины главных максимумов меньше полуширины полос интерферометра с бóльшим расстоянием между пластинами. Практически a берется равным 3 или большему числу, и из (130) видно, что увеличение F_2 вызывает значительно большее уменьшение ε_2, чем такое же увеличение F_1. Следует также помнить, что максимальное пропускание такой комбинации интерферометров равно произведению $\tau_1\tau_2$, т. е. величин максимального пропускания обоих интерферометров, как это следует из (122) и (33); в самом деле, мы видели в п. 7.6.2, что при использовании соответствующих отражающих покрытий увеличение $\mathcal{R}_1$ и $\mathcal{R}_2$ (а следовательно, F_1 и F_2) сопровождается уменьшением τ_1 и τ_2. Поэтому не следует делать F_1 больше, чем это необходимо для подавления вторичных максимумов до приемлемой величины, и наоборот, F_2 нужно делать как можно больше, пока уменьшение интенсивности не поставит этому предел. При таких условиях F_1/a^2 столь мало по сравнению с F_2, что $\varepsilon_2 \approx 4/\sqrt{F_2}$ и разрешающая сила подобной комбинации интерферометров почти равна разрешающей силе интерферометра с бóльшим расстоянием между пластинами. Присутствие вторичных максимумов неудобно, если спектральные компоненты исследуемого источника сильно различаются по интенсивностям, так как главные максимумы слабой компоненты можно принять за вторичные максимумы сильной компоненты. Такая неопределенность устраняется, если наблюдения ведутся с комбинациями интерферометров с различными расстояниями между пластинами.

Если $a = 1$, вторичные максимумы отсутствуют, и из (122) и (34) следует, что контрастность, достигаемая с комбинацией интерферометров, равна произведению $\mathcal{C}_1\mathcal{C}_2$ контрастностей обоих интерферометров. Для данного максимального пропускания с такой комбинацией интерферометров достигается значительно более высокая контрастность, чем с одиночным интерферометром; поэтому подобное устройство особенно ценно при наблюдении слабых сателлитов спектральных линий [50, 91].

б. Полосы суперпозиции. Большой практический интерес представляют полосы, получающиеся с двумя плоскопараллельными пластинками, наклоненными под углом α друг к другу и освещенными светом, падающим почти нормально. Предполагается, что падающий свет столь сильно отличается от монохроматического, что с каждой отдельной пластинкой интерференционные полосы не наблюдаются. Чтобы разобраться во всех обстоятельствах, связанных с этим *), рассмотрим сначала прохождение монохроматической волны с волновым числом $k_0 = 2\pi/\lambda_0$ сквозь одну пластинку. Пусть $A^{(i)}(k_0)$ и $A^{(t)}(k_0)$ — комплексные амплитуды соответственно падающей и прошедшей волн. Принимая во внимание все отражения, получим, согласно (10) и учитывая (2) и (4),

$$A^{(t)}(k_0) = A^{(i)}(k_0)\mathcal{T} \sum_{p=0}^{\infty} \mathcal{R}^p \exp(ip\delta), \tag{131}$$

а интенсивность $I^{(t)} = A^{(t)}A^{(t)*}$ прошедшего света выразится соотношением

$$I^{(t)}(k_0) = I^{(i)}(k_0)\mathcal{T}^2 \sum_{p=0}^{\infty} \sum_{p'=0}^{\infty} \mathcal{R}^p\mathcal{R}^{p'} \exp[i(p-p')\delta], \tag{132}$$

где $I^{(i)} = A^{(i)}A^{(i)*}$ — интенсивность падающего света. Если мы примем

*) Приведенный здесь анализ в основном принадлежит Бенуа, Фабри и Перо [92].

$|p-p'|=q$, то (132) можно переписать в виде

$$I^{(t)}(k_0)=I^{(i)}(k_0)\mathcal{T}^2\sum_{p=0}^{\infty}\mathcal{R}^{2p}\left\{1+\sum_{q=1}^{\infty}\mathcal{R}^q(\exp(iq\delta)+\exp(-iq\delta)\right\}=$$

$$=I^{(i)}(k_0)\frac{\mathcal{T}^2}{1-\mathcal{R}^2}\left(1+2\sum_{q=1}^{\infty}\mathcal{R}^q\cos q\delta\right). \quad (133)$$

Интенсивность монохроматического света, прошедшего сквозь две пластинки, поставленные последовательно (если пренебречь светом, отраженным в прямом и обратном направлениях между двумя пластинами), определяется выражением

$$I^{(t)}(k_0)=$$

$$=I^{(i)}(k_0)\frac{\mathcal{T}_1^2\mathcal{T}_2^2}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)}\left(1+2\sum_{r=1}^{\infty}\mathcal{R}_1^r\cos r\delta_1\right)\times\left(1+2\sum_{s=1}^{\infty}\mathcal{R}_2^s\cos s\delta_2\right), \quad (134)$$

где индексы 1 и 2 относятся соответственно к первой и второй пластинке. Разности фаз δ_1 *и* δ_2 *находят из соотношений*

$$\left.\begin{aligned}\delta_1&=k_0\,\Delta\mathcal{S}_1+2\varphi_1=k_0 2n_1'h_1\cos\theta_1'+2\varphi_1,\\ \delta_2&=k_0\,\Delta\mathcal{S}_2+2\varphi_2=k_0 2n_2'h_2\cos\theta_2'+2\varphi_2,\end{aligned}\right\} \quad (135)$$

где разные символы имеют то же значение, что и раньше.

Однако если свет не монохроматичен, мы можем рассматривать его как суперпозицию монохроматических компонент с различными частотами. Согласно п. 7.5.8 отдельные компоненты некогерентны и результирующая интенсивность равна сумме (интегралу) интенсивностей отдельных компонент. Таким образом, из (134) получается следующее выражение для полной интенсивности света, прошедшего через обе пластинки:

$$I^{(t)}=\int\frac{\mathcal{T}_1^2\mathcal{T}_2^2}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)}\,i^{(i)}(k_0)\left(1+2\sum_{r=1}^{\infty}\mathcal{R}_1^r\cos r\delta_1\right)\left(1+2\sum_{s=1}^{\infty}\mathcal{R}_2^s\cos s\delta_2\right)dk_0, \quad (136)$$

где $i^{(i)}(k_0)$ — спектральное распределение интенсивности падающего света. Величины $\mathcal{R}$ и $\mathcal{T}$, вообще говоря, являются функциями k_0, но мы допустим, что в области, где $i^{(i)}(k_0)$ значительно, их изменения в зависимости от k_0 ничтожны. Первый множитель в (136) можно тогда вынести за знак интеграла, и окончательно это выражение можно переписать в виде

$$I^{(t)}=\frac{\mathcal{T}_1^2\mathcal{T}_2^2}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)}\int i^{(i)}(k_0)\{1+2(\sigma_1+\sigma_2+\sigma_{12}^{+}+\sigma_{12}^{-})\}\,dk_0, \quad (137а)$$

где

$$\left.\begin{aligned}\sigma_1&=\sum_{r=1}^{\infty}\mathcal{R}_1^r\cos r\delta_1, \qquad \sigma_2=\sum_{s=1}^{\infty}\mathcal{R}_2^s\cos s\delta_2,\\ \sigma_{12}^{+}&=\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\mathcal{R}_1^r\mathcal{R}_2^s\cos(r\delta_1+s\delta_2), \qquad \sigma_{12}^{-}=\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\mathcal{R}_1^r\mathcal{R}_2^s\cos(r\delta_1-s\delta_2).\end{aligned}\right\} \quad (137б)$$

Если монохроматические компоненты занимают область длин волн $\Delta\lambda_0$ вблизи средней длины волны $\bar{\lambda}_0$, то соответствующая область k_0 равна $2\pi\Delta\lambda_0/\bar{\lambda}_0^2$ и при $\Delta\mathcal{S}_1$ и $\Delta\mathcal{S}_2$, достаточно больших по сравнению с длиной когерентности $\bar{\lambda}_0^2/\Delta\lambda$, области изменения δ_1 и δ_2 велики по сравнению *) с 2π. При таких усло-

*) Интегрирование выражения (133) по всем спектральным компонентам ясно показывает, что в любой отдельной пластинке интенсивность прошедшего света практически не зависит от δ, т. е. полосы отсутствуют.

виях σ_1, σ_2 и σ_{12}^+ быстро изменяются в области интегрирования и многократно изменяют знаки. Следовательно, вклад этих членов в $I^{(t)}$ невелик, и (137) сводится к *)

$$I^{(t)} \approx \frac{\mathcal{T}_1^2\mathcal{T}_2^2}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)} \int i^{(i)}(k_0)\left(1+2\sum_{r=1}^{\infty}\sum_{s=1}^{\infty} \mathcal{R}_1^r\mathcal{R}_2^s \cos(r\delta_1 - s\delta_2)\right) dk_0. \qquad (138)$$

В общем случае величины $r\delta_1 - s\delta_2$ по крайней мере такого же порядка, как и δ_1 или δ_2, а, значит, интегралы от членов с косинусом в (138) также незначительны, и поэтому $I^{(t)}$ действительно не зависит от δ_1 и δ_2. Существует, однако, исключение, когда

$$a\left(\Delta\mathcal{S}_1+\frac{2\varphi_1}{k_0}\right)-b\left(\Delta\mathcal{S}_2+\frac{2\varphi_2}{k_0}\right)=\varepsilon, \qquad (139)$$

где a и b — небольшие целые числа, не содержащие общего множителя, а $|\varepsilon|$ невелико по сравнению с $\bar{\lambda}_0^2/\Delta\lambda_0$. В этом случае при

$$r=qa, \qquad s=qb \qquad (q=1,\ 2,\ 3,\ \ldots) \qquad (140)$$

получим

$$r\delta_1 - s\delta_2 = qk_0\varepsilon, \qquad (141)$$

и область значений $|r\delta_1 - s\delta_2|$ равна $|q\varepsilon\Delta k_0|$, что невелико по сравнению с $2\pi q$; при таких значениях r и s интегралы членов, содержащих косинусы, в (138) не обязательно пренебрежимо малы, и мы получим

$$I^{(t)}(\varepsilon) \approx \frac{\mathcal{T}_1^2\mathcal{T}_2^2}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)} \int i^{(i)}(k_0)\left(1+2\sum_{q=1}^{\infty}(\mathcal{R}_1^a\mathcal{R}_2^b)^q \cos qk_0\varepsilon\right) dk_0. \qquad (142)$$

Ряд под знаком интеграла в (142) идентичен ряду в выражении (133) для интенсивности волны монохроматического света, прошедшей сквозь одиночную пластину. В п. 7.6.1 такой ряд был просуммирован (см. уравнение (13)). Следовательно, сумму ряда в (142) можно написать сразу в виде

$$1+2\sum_{q=1}^{\infty}(\mathcal{R}_1^a\mathcal{R}_2^b)^q \cos qk_0\varepsilon = \frac{1-(\mathcal{R}_1^a\mathcal{R}_2^b)^2}{(1-\mathcal{R}_1^a\mathcal{R}_2^b)^2+4\mathcal{R}_1^a\mathcal{R}_2^b \sin^2\frac{k_0\varepsilon}{2}}. \qquad (143)$$

Из (142) и (143) находим

$$I^{(t)}(\varepsilon) = \frac{\mathcal{T}_1^2\mathcal{T}_2^2\{1-(\mathcal{R}_1^a\mathcal{R}_2^b)^2\}}{(1-\mathcal{R}_1^2)(1-\mathcal{R}_2^2)(1-\mathcal{R}_1^a\mathcal{R}_2^b)^2} \int \frac{i^{(i)}(k_0)}{1+B\sin^2\left(\frac{k_0\varepsilon}{2}\right)} dk_0, \qquad (144)$$

где

$$B = \frac{4\mathcal{R}_1^a\mathcal{R}_2^b}{(1-\mathcal{R}_1^a\mathcal{R}_2^b)^2}. \qquad (145)$$

Сравнивая с (15б), мы видим, что распределение интенсивности (144) эквивалентно суперпозиции распределений интенсивности монохроматических компонент. Форма распределения для каждой компоненты совпадает с показанными на рис. 7.58. Максимумы интенсивности в этих распределениях появляются при $k_0\varepsilon$, равному целому кратному 2π, т. е. при

$$\varepsilon = m\lambda_0, \qquad |m| = 0,\ 1,\ 2,\ \ldots \qquad (146)$$

Если предположить, что n_1', n_2', φ_1/k_0 и φ_2/k_0 постоянны в спектральной области, включенной в (144), то максимумы нулевого порядка ($\varepsilon=0$) в распределениях интенсивности для монохроматических компонент совпадают в цент-

*) Строгое доказательство возможности перехода от (137а) к (142) требует более точного рассмотрения, чем это сделано здесь.

ральной полосе. Согласно (135) и (139) ее положение в фокальной плоскости линзы L (см. рис. 7.80) определяется соотношением

$$\frac{n_2'h_2\cos\theta_2'+c_2}{n_1'h_1\cos\theta_1'+c_1}=\frac{a}{b}, \tag{147}$$

где c_1 и c_2 обозначают φ_1/k_0 и φ_2/k_0 соответственно. При увеличении $|\varepsilon|$ с обеих сторон этой полосы интерференционные картины от компонент взаимно смещаются, так как размер каждой картины пропорционален соответствующей ей длине волны, и поэтому четкость полос уменьшается. В белом свете видна белая центральная полоса, положение которой определяется (147), окруженная с каждой стороны цветными максимумами и минимумами. По мере удаления от центра освещенность для глаза приближается к равномерной. Такие полосы называются *полосами суперпозиции*. Они представляют собой многолучевой вариант полос Брюстера (см. п. 7.5.6) и, подобно последним, имеют вид прямых линий, параллельных ребру клина, образованного пластинами. Расстояние между этими полосами обратно пропорционально углу клина α.

Полосы суперпозиции можно использовать, как предложили Фабри и Бюиссон [93], для определения разности оптических толщин двух эталонов Фабри — Перо, отношение оптических толщин которых очень близко к целому числу a. Для этой цели один эталон укрепляют неподвижно, а другой наклоняют по отношению к первому до тех пор, пока центральная белая полоса его интерференционной картины не пройдет через точку O, т. е. фокус линзы L для света, прошедшего нормально через неподвижный эталон (см. рис. 7.80). При фиксированном первом эталоне $\theta_1=\theta_1'=0$ и $\theta_2=\alpha$, где α — угол между эталонами. Для малого α по закону преломления имеем $\theta_2'\approx n\alpha/n_2'$, где n — показатель преломления воздуха, окружающего эталоны; $\cos\theta_2'\approx 1-\theta_2'^2/2\approx 1-n^2\alpha^2/n_2'^2$ и из (147), принимая $b=1$, получим с точностью до второй степени α

$$(n_2'h_2+c_2)-a(n_1'h_1+c_1)=\frac{n^2h_2}{2n_2'}\alpha^2. \tag{148а}$$

Если же второй эталон неподвижен, то белая полоса проходит через точку O, когда

$$a(n_1'h_1+c_1)-(n_2'h_2+c_2)=\frac{n^2ah_1}{2n_1'}\alpha^2. \tag{148б}$$

Если эталоны эвакуированы (что иногда делается), то можно принять $n_1'=n_2'=n$ при условии, что разность $|a(n_1'h_1+c_1)-(n_2'h_2+c_2)|$ достаточно мала; тогда измерения α дают эту разность, выраженную через оптическую толщину одного из эталонов. Таким способом были выполнены измерения при значениях a, равных 10.

Сходное устройство, только с эталонами приблизительно равной толщины, было использовано для измерений показателя преломления и дисперсии воздуха [94]. Сначала эталоны эквакуировали и наклоном одного из них устанавливали удобный интервал между полосами в почти монохроматическом свете со средней длиной волны $\bar{\lambda}_0$. Затем заполняли воздухом неподвижный эталон толщиной h, что приводило к изменению ε на $2(n-1)h$, где n — показатель преломления воздуха. Подсчитывая соответствующее число Δm полос, прошедших через O, имеем в согласии с (146)

$$\Delta m=\frac{\Delta\varepsilon}{\bar{\lambda}_0}=\frac{2(n-1)h}{\bar{\lambda}_0}; \tag{149}$$

отсюда, если известны h и λ_0, находим $(n-1)$,

§ 7.7. Сравнение длин волн с эталонным метром

Эталон длины представляет собой расстояние между двумя штрихами, выгравированными на стержне из сплава платина — иридий, при температуре 0 °С. Этот стержень, хранящийся во Франции, называется *международным эталоном метра.* Как уже упоминалось в п. 7.6.4, соотношение между оптическими длинами волн и фактическими длинами основано на сравнении длины волны красной линии (6438 Å) кадмия с длиной этого эталона. Первое сравнение было выполнено в 1892 г. Майкельсоном и Бенуа [26] с помощью видоизмененного интерферометра Майкельсона. В 1905 г. такое измерение было повторено с большей точностью Бенуа, Фабри и Перо [92]. Они воспользовались пятью эталонами Фабри — Перо с длинами, примерно равными 6,25; 12,5; 25; 50 и 100 *см.* Число длин волн красной кадмиевой линии, содержащихся в длине самого короткого эталона, определялось методом дробных частей порядка, который требует знания достаточно точного отношения длин волн (см. п. 7.6.4), а не самой их длины. Длина каждого эталона сравнивалась затем с длиной следующего по величине эталона с помощью интерференционных полос в белом свете. Для этого два эталона устанавливались параллельно и освещались белым светом. Прошедший свет освещал тонкий воздушный клин, образованный двумя полупрозрачными посеребренными плоскими поверхностями. При этих условиях возникали полосы, локализованные в клине, подобные описанным в п. 7.6.8. *б.* Центральная полоса располагалась вдоль линии, где толщина клина равнялась разности между длиной большего эталона и удвоенной длиной более короткого эталона. Эта разность была определена в длинах волн красной кадмиевой линии путем предварительной градуировки толщины клина. После четырех последовательных сравнений эталонов между собой было определено число длин волн, укладывающихся в эталоне длиной 100 *см.* Наконец, определялась разница между длиной этого эталона и длиной копии эталона метра. Эта часть эксперимента включала установку особо точных подвижных измерительных микроскопов на градуировочные штрихи метра и на такие же градуировочные штрихи, нанесенные на боковые поверхности пластин эталона. Расстояние между последними штрихами и отражающими поверхностями эталона определялось дополнительными опытами. Окончательный результат, полученный для длины волны красной линии кадмия в сухом воздухе при 15 °С и при давлении 760 *мм рт. ст.* равнялся $6438{,}4696 \cdot 10^{-10}$ *м.* Возможная относительная ошибка в интерферометрических измерениях составляла около $1 \cdot 10^{-7}$.

Повторные определения (библиографию см. в [95]) соотношения между длиной волны красной линии кадмия и метром были выполнены в целом ряде лабораторий по стандартизации таким же методом, как и метод Бенуа, Фабри и Перо или в принципе схожим с ним. Интересны эксперименты Сирса и Баррелла [96], так как в них сделаны прямые измерения длины волны в вакууме. Они воспользовались только тремя эталонами Фабри — Перо; самый большой имел в длину немного больше метра, другие — приблизительно одну треть и одну девятую метра. Разделителями служили цилиндры из инвара с оптически плоскими хромированными торцами, к которым прижимались эталонные пластины. Эти соединения были герметизированы, и эталоны можно было эвакуировать. Было измерено число длин волн, укладывающееся в самом коротком эталоне. Измерения делались методом дробных частей порядка, а для сравнения эталонов друг с другом применялись полосы суперпозиции, описанные в п. 7.6.8. Самый большой эталон имел достаточную длину, чтобы в нем могла поместиться стальная концевая мера номинальной длиной в 1 *м.* Расстояние в длинах волн между полированными торцами концевой меры и отражающими поверхностями эталона определялось путем наблюдения в отраженном свете полос, локализованных в бесконечности. Таким способом была

определена длина концевой меры в длинах волн красной кадмиевой линии с точностью около $2 \cdot 10^{-8}$. Окончательно концевая мера сравнивалась с копией метра хорошо разработанным методом сравнения концевых эталонов с штриховыми *).

После того как результаты всех измерений были приведены к длине волны в воздухе при нормальных условиях (воздух при 15 °С и давлении 760 *мм рт. ст.*, свободный от паров воды и содержащий 0,03% углекислого газа), было найдено среднее значение $6438{,}4696 \cdot 10^{-10}$ *м*, которое случайно совпало с величиной, принятой в качестве определения ангстрема. Наибольшее отклонение в любом отдельном измерении от среднего составляло $3 \cdot 10^{-5}$% от измеренной величины или около $3 \cdot 10^{-4}$ *мм* на 1 *м*. Это значительно меньше, чем ширина градуировочных штрихов на эталоне метра и на его копиях, использованных в экспериментах. Несомненно, что это внесло неопределенность в измерения, сделанные с помощью микроскопа. Интерферометрические измерения можно осуществлять со значительно большей точностью, и поэтому естественно предусмотреть замену материального эталона метра определением длины метра в длинах волн какой-то спектральной линии.

Международное бюро мер и весов приняло в 1954 г. рекомендацию по этому вопросу [98] с оговоркой, что для сохранения преемственности эталона длины его новые определения должны быть согласованы с величиной $6438{,}4696 \cdot 10^{-10}$ *м* для длины волны красной линии кадмия. В 1958 г. Международное бюро постановило [99], что лучше всего для этой цели подходит спектральная линия с длиной волны приблизительно 6056 Å, соответствующая переходу между уровнями $2p_{10}$ и $5d_5$ атома криптона с массовым числом 86; на основании измерений, сделанных в пяти различных лабораториях, бюро предложило, чтобы метр был определен точно в 1 650 763,73 длины волны в вакууме этого излучения. Такое определение было одобрено в 1960 г. 11-й Генеральной конференцией по мерам и весам [100]). Следовательно, первичные эталоны длин волн и первичный эталон длины теперь идентичны и ангстрем точно равен 10^{-10} *м*.

ЛИТЕРАТУРА

1. A. A. Michelson, Astrophys. J. **47**, 283 (1918).
2. Rayleigh, Proc. Roy. Soc. **59**, 198 (1896); F. Haber, F. Löwe, Z. angew. Chem. **23**, 1393 (1910).
3. A. A. Michelson F. G. Pease, Astrophys. J. **53**, 256 (1921).
4. H. Fizeau, C. R. Akad Sci **66**, 934 (1868).
5. A. A. Michelson, Phil Mag **30**, 1 (1890).
6. A. A. Michelson, Nature **45**, 160 (1891).
7. J. A. Anderson, Astrophys. J. **51**, 263 (1920).
8. A. A. Michelson, Astrophys. J. **51**, 257 (1920); A. A. Michelson, F. G. Pease, Astrophys. J. **53**, 249 (1921).
9. J. L. Pawsey, R. N. Bracewell, Radio Astronomy, Clarendon Press, Oxford, 1955 R. N. Bracewell, в сб. «Encyclopedia of Physics», ed. S. Flügge, Springer, Berlin,; 1959, Vol. 54.
10. R. H. Brown, R. Q. Twiss, Phil. Mar. **45**, 663 (1954).
11. R. H. Brown, R. Q. Twiss, Nature **178**, 1046 (1956); Proc. Roy. Soc. **A248**, 199, 222 (1958)
12. O. Wiener, Ann. d. Physik **40**, 203 (1890).
13. P. Drude, W. Nernst, Wiedem. Ann **45**, 460 (1892)
14. H. E. Ives, T. C. Fry, J. Opt. Soc. Amer. **23**, 73 (1933).
15. G. Lippmann, C. R. Acad. Sci. **112**, 274 (1891).
16. F. Twyman, Prism and Lens Making, Hilger a. Watts, Ltd, London 2nd ed., 1952, p. 388.
17. H. Fizeau, Ann. Chim. Phys. **66**, 429 (1862)
18. F. H. Rolt, Engineering **144**, 162 (1937)
19. J. R. Benoit, J. de Phys. **7**, 57 (1898).
20. J. Macé, de Lépinaw, C. Fabry, J. de Phys. **10**, 5 (1891).

*) Более детально см. [97].

21. J. Macé de Lépinay, J. de Phys. 9, 121 (1890); W. Feussner, L. Janicki, в сб. «Handbuch der physikalischen Optik», ed. E. Gehrcke, Barth, Leipzig, 1927, Vol. I., p. 396; G. F. C. Searle, Phil. Mag. 37, 361 (1946).
22. A. A. Michelson, Amer. J. Sci. 22, 120 (1881); Phil. Mag. 13, 236 (1882).
23. R. J. Kennedy, Proc. Nat. Acad. Sci. 12, 621 (1926).
24. K. K. Illingworth, Phys. Rev. 30, 692 (1927).
25. A. A. Michelson, E. W. Morley, Phil. Mag. 24, 449 (1887).
26. A. A. Michelson, J. R. Benoit, Trav. et Mem. Int. Bur. Poids et Mes. 11, 1 (1895).
27. F. Twyman, A. Green, British Patent № 103832 (1916).
28. E. Twyman, Phil. Mag. 35, 49 (1918).
29. F. Twyman, Phil. Mag. 42, 777 (1921).
30. R. Kingslake, Trans. Opt. Soc. 27, 94 (1927).
31. W. Kösters, в сб. «Handbuch der Physikalischen Optik», ed. E. Gehrcke, Barth, Leipzig, 1927, Vol. I, p. 484.
32. J. H. Dowell, British Patent № 555672 (1942).
33. J. Jamin, C. R. Akad. Sci. 42, 482 (1856).
34. J. A. Sirks, Hd. Ned. Nat. en Geneesk. Congr., Groningen, 1893, p. 92.
35. E. Pringsheim, Verh. Phys. Ges. Berlin, 17, 152 (1898).
36. J. Dyson. Proc. Roy. Soc. A204, 170 (1950).
37. L. Zehnder, Z. f. Instrkde 11, 275 (1891); L. Mach, Z. f. Instrkde 12, 89 (1892).
38. R. Ladenburg, D. Bershader, в сб. «Interferometry in High Speed Aerodynamics and Jet Propulsion», Vol. IX, Physical Measurements in Gas Dynamics and Combustion, Oxford Univ. Press, 1955, Art. A. 3.
39. W. J. Bats, Proc. Phys. Soc. 59, 940 (1947).
40. R. L. Drew, Proc. Phys. Soc. B64, 1005 (1951).
40a. O. Brungdahl, в сб.: «Progr. in Optics», vol. 4, ed. E. Wolf. North Holland Publ. Co., Amsterdam, a. New York, J. Wiley a. Sons, N.Y., 1965, p. 37.
41. S. O. Rice, Bell. Tech. J. 23, 282 (1944).
42. A. A. Michelson, Phil. Mag. 31, 388 (1891).
43. Rayleigh, Phil. Mag. 34, 407 (1892).
44. E. Wolf, Proc. Phys. Soc. 80, 1269 (1962).
45. A. Pérard, Rev. d'Opt. 7, 1 (1928); Réunions l'Institut d'Optique Rev. d'Opt. 6me année, 10 (1935).
46. A. A. Michelson. Phil. Mag. 34, 291 (1892).
47. P. Fellgett, J. de Phys. 19, 187, 237 (1958).
47a. G. A. Vanass, H. Sakai, в сб.: «Progr. in Optics», vol. 6, ed. E. Wolf. North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1967, p. 259.
48. P. Giacomo, Rev. d'Opt. 35, 442 (1956).
49. C. Fabry, A. Perot, Ann. Chim. Phys. 16, 115 (1899).
50. C. Dufour, Ann. de Physique 6, 5 (1951).
51. P. Connes, Rev. d'Opt. 35, 37 (1956); J. de Phys. 19, 262 (1958).
52. A. G. Fax, T. Li, Bell. Tech. J. 40, 453 (1961).
53. G. D. Boyd, J. P. Gordon, Bell. Tech. J. 40, 489 (1961).
54. C. Dufour, R. Picca, Rev. d'Opt, 24, 19 (1945).
55. R. Chabbal, J. Rech. Cent. Nat. Rech. Sci. № 24, 138 (1953).
56. H. Kuhn, B. A. Wilson, Proc. Phys. Soc. B63, 745 (1950).
57. J. C. Burridge, H. Kuhn, A. Pery, Proc. Phys. Soc. B66, 963 (1953).
58. P. Giacomo, Rev. d'Opt. 35, 317 (1956); J. Ring, W. L. Wilcock, Nature 171, 648 (1953); 173, 994 (1954).
59. S. Penselin, A. Steudel, Z. Phys. 142, 21 (1955).
60. Rayleigh, Phil. Mag. 8, 261 (1879).
61. S. Tolansky, High Resolution Spectroscopy, Methuen, London, 1947, Ch. 9. (С. Толанский, Спектроскопия высокой разрешающей силы, ИЛ, 1955, гл. 9.)
62. C. Fabry, H. Buisson, J. de Phys. 9, 197 (1910).
63. P. Jacquinot, C. Dufour, J. Rech. Cent. Nat. Rech. Sci. № 6, 91 (1948).
64. P. Jacquinot, J. Opt. Soc. Amer. 44, 761 (1954).
65. Trans. Int. Union for Co-operation in Solar Research, Manchester Univ. Press, 1908, Vol. 2, p. 17.
66. Proc. Verb. Com. Int. Poids et Mes. 17, 91 (1935).
67. C. Fabry, A. Perot, Ann. Chim. Phys. 16, 289 (1899).
68. C. V. Jackson, Phil. Trans. Roy. Soc. A236, 1 (1936).
69. G. R. Harrison, J. Opt. Soc. Amer. 36, 644 (1946).
70. K. W. Meissner, J. Opt. Soc. Amer. 31, 416 (1941).
71. K. W. Meissner, J. Opt. Soc. Amer. 32, 191 (1942).
72. H. D. Babcock, Astrophys. J. 65, 140 (1927).
73. O. Lummer, Verh. Deutsch. Phys. Ges. 3, 85 (1901); O. Lummer, E. Gehrhcke, Ann. d. Physik 10, 457 (1903).
74. C. Dufour. Rev. d'Opt. 24, 11 (1945).
75. C. Dufour, Rev. d'Opt. 31, 1 (1952).

76. A. F. Turner, J. de Phys. 11, 457 (1950).
77. P. H. Lissberger, J. Ring, Optica Acta 2, 45 (1955).
78. K. M. Greenland, C. Billington, J. de Phys. 11, 418 (1950).
79. P. Giacomo, P. Jacquinot, J. de Phys. 13, 59A (1952).
80. P. Lissberger, J Ring, Optica Acta 2, 42 (1955).
81. P. Leurgans, A. F. Turner, J. Opt. Soc. Amer. 37, 983 (1947).
82. A. F. Turner, J. de Phys. 11, 458 (1950)
83. B. H. Billings, J Opt. Soc. Amer. 39, 634 (1949).
84. J. Brossel, Proc. Phys. Soc. 59, 224 (1947)
85. K. Kinosita, J. Phys Soc. Japan 8, 219 (1953).
86. J. Holden, Proc. Phys. Soc. B62, 405 (1949).
87. S. Tolansky, Multiple-beam Interferometry of Surfaces and Films, Clarendon Press, Oxford, 1948.
88. S. Tolansky, Phil. Mag. 36, 225 (1945).
89. S. Tolansky, W. L Wilcock, Proc. Roy. Soc A191, 192 (1947).
90. W. V. Houston, Phys. Rev. 29, 478 (1927); E. Gehrcke, E. Lau, Z. tech. Physik 8, 157 (1927).
91. L. C. Bradley, H. Kuhn, Nature 162, 412 (1948).
92. J. R. Benoit, C. Fabry, A. Perot, Trav. et Mem. Bur. Int. Poids et Mes. 15, 1 (1913).
93. C. Fabry, H. Buisson, J. de Phys. 9, 189 (1919).
94. H. Barrell, J. E. Sears, Phil. Trans Roy Soc. A238, 1 (1939).
95. H. Barrell, Proc. Roy. Soc. A186, 164 (1946)
96. J. E. Sears, H. Barrell, Phil. Trans. Roy. Soc. A231, 75 (1932); A233, 143 (1934).
97. J. E. Sears, Proc. Roy. Soc. A186, 152 (1946)
98. Proc. Verb. Com. Int Poids et Mes. 24, 2 (1954)
99. Proc. Verb. Com. Int. Poids et Mes. (2), 26-B, M 30 (1958).
100. C. R. 11-me Conf. Gén. Poids et Mes., Gauthier-Villars, 1960, pp. 51, 85.

Г Л А В А 8

ЭЛЕМЕНТЫ ТЕОРИИ ДИФРАКЦИИ

§ 8.1. Введение

Переходя от общего электромагнитного поля к оптическому полю, характеризующемуся очень высокими частотами (короткие волны), мы нашли, что в определенных областях простая геометрическая модель распространения энергии становится неверной. В частности, мы увидели, что отклонений от нее нужно ожидать в непосредственной близости к границам тени и в областях, где концентрируется большое число лучей. Подобные отклонения проявляются в появлении темных и светлых линий — дифракционных полос. Теория дифракции занимается главным образом изучением поля в этих особых областях, и естественно, что такие области, в которых лежит часть пространства, где образуется оптическое изображение (область фокуса), представляют большой практический интерес.

Первое упоминание о дифракционных явлениях появилось в работе Леонардо да Винчи (1452—1519 гг.). Однако впервые они были описаны детально только в книге Гримальди, опубликованной в 1665 г. спустя два года после его смерти. Корпускулярная теория, которую считали в то время правильно описывающей распространение света, не могла объяснить дифракцию. Гюйгенс, впервые обосновавший волновую теорию, очевидно, не знал об открытии Гримальди, иначе он несомненно сослался бы на него для подтверждения своей точки зрения. О возможности объяснить явления дифракции в рамках волновой теории нигде не упоминается вплоть до 1818 г., когда появился прекрасный мемуар Френеля (см. «Историческое введение»), где было показано, что явление дифракции можно объяснить с помощью построения Гюйгенса (см. п. 3.3.3) и применения принципа интерференции. Позднее Кирхгоф (1882 г.) придал исследованиям Френеля строго математическое обоснование, и с этого времени началось широкое изучение дифракции *).

Проблемы, возникающие при изучении дифракционных явлений, относятся к наиболее трудным в оптике, и их редко удается довести до строгого решения. Первое такое решение было получено только в 1896 г. А. Зоммерфельдом, рассмотревшим важный вопрос дифракции плоской волны на идеально проводящем полубесконечном плоском экране. С тех пор было найдено строгое решение только нескольких дифракционных задач, относящихся главным образом к двумерным структурам (см. гл. 11). В большинстве же случаев, представляющих практический интерес, из-за математических трудностей приходится прибегать к приближенным методам, и тут теория Гюйгенса и Френеля служит чрезвычайно мощным орудием, позволяющим решить большинство вопросов, встречающихся в инструментальной оптике. Эта теория и некоторые ее приложения составляют главное содержание настоящей главы.

§ 8.2. Принцип Гюйгенса — Френеля

Согласно построению Гюйгенса (см. п. 3.3.3) каждую точку волнового фронта можно считать центром вторичного возмущения, которое вызывает элементарные сферические волны, а волновой фронт в любой более поздний момент времени — огибающей этих волн. Френель смог объяснить явление

*) Более полное знакомство с историческим развитием этого предмета можно получить в [1].

дифракции, дополнив построение Гюйгенса утверждением, что вторичные волны интерферируют между собой. Это сочетание построения Гюйгенса с принципом интерференции называется *принципом Гюйгенса — Френеля*. Прежде чем применить его к изучению дифракционных эффектов, следует проверить (с некоторыми простыми дополнительными предположениями), правильно ли он описывает распространение света в вакууме.

Пусть S (рис. 8.1) — мгновенное положение сферического монохроматического волнового фронта с радиусом r_0, распространяющегося от точечного источника P_0; допустим, что требуется определить световое возмущение в

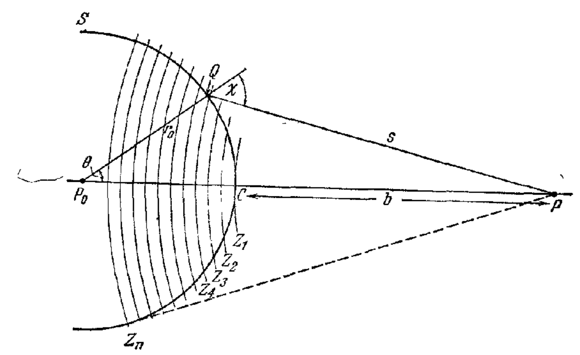


Рис. 8.1. Построение зон Френеля.

точке P. Возмущение в точке Q волнового фронта можно представить (с точностью до периодического по времени множителя $\exp(-i\omega t)$) в виде $A \exp(ikr_0)/r_0$, где A — амплитуда на расстоянии единицы длины от источника. В соответствии с принципом Гюйгенса — Френеля каждый элемент волнового фронта рассматривается как центр вторичных возмущений, которые распространяются в виде элементарных сферических волн; вклад в возмущение $dU(P)$, вносимый элементом dS, находящимся в точке Q, запишется в виде

$$dU(P) = K(\chi) \frac{A \exp(ikr_0)}{r_0} \frac{\exp(iks)}{s} dS,$$

где $s = QP$, $K(\chi)$ — *коэффициент наклона*, описывающий изменение амплитуды вторичных волн в зависимости от направления, а χ — угол (часто называемый *углом дифракции*) между нормалью в точке Q и направлением QP. Следуя Френелю, предположим, что K максимально в первоначальном направлении распространения света, т. е. при $\chi = 0$, быстро уменьшается с увеличением χ и равно нулю, когда QP становится касательной к волновому фронту, т. е. когда $\chi = \pi/2$; наконец, предположим, что в точке P сказывается влияние только той части S' первоначальной волны, которая не загораживается каким-либо препятствием, находящимся между P_0 и P. Поэтому полное возмущение в точке P равно

$$U(P) = \frac{A \exp(ikr_0)}{r_0} \iint_{S'} \frac{\exp(iks)}{s} K(\chi)\, dS. \qquad (1)$$

Для того чтобы вычислить (1), воспользуемся так называемыми *зонами Френеля*. Построим вокруг точки P сферы с радиусами

$$b,\ b+\frac{\lambda}{2},\ \ b+\frac{2\lambda}{2},\ \ b+\frac{3\lambda}{2},\ \ldots,\ b+\frac{j\lambda}{2},\ \ldots,$$

где $b = CP$, а C — точка пересечения P_0P с волновым фронтом S (см. рис. 8.1). Сферы делят S на целый ряд зон $Z_1, Z_2, Z_3, \ldots, Z_j, \ldots$

Пусть r_0 и b велики по сравнению с длиной волны; тогда можно предположить, что в любой зоне величина K постоянна и в зоне j равна K_j. Из рисунка видно, что

$$s^2 = r_0^2 + (r_0 + b)^2 - 2r_0 (r_0 + b) \cos\theta.$$

Следовательно,

$$s\,ds = r_0 (r_0 + b) \sin\theta\, d\theta \tag{2}$$

и, значит,

$$dS = r_0^2 \sin\theta\, d\theta\, d\varphi = \frac{r_0}{r_0 + b}\, s\, ds\, d\varphi,$$

где φ — азимут. Следовательно, вклад зоны j в $U(P)$ равен

$$U_j(P) = 2\pi\, \frac{A \exp(ikr_0)}{r_0 + b} K_j \int\limits_{b+(j-1)\lambda/2}^{b+j\lambda/2} \exp(iks)\, ds =$$

$$= -\frac{2\pi i}{k} K_j \frac{A \exp[ik(r_0 + b)]}{r_0 + b} \exp\left(\frac{ikj\lambda}{2}\right)\left(1 - \exp\left(-\frac{ik\lambda}{2}\right)\right).$$

Так как $k\lambda = 2\pi$, последние два множителя сводятся к

$$\exp\left(\frac{ikj\lambda}{2}\right)\left(1 - \exp\left(-\frac{ik\lambda}{2}\right)\right) = \exp(i\pi j)\,(1 - \exp(-i\pi)) = (-1)^j \cdot 2$$

и, следовательно,

$$U_j(P) = 2i\lambda\, (-1)^{j+1} K_j \frac{A \exp[ik(r_0 + b)]}{r_0 + b}. \tag{3}$$

Заметим, что вклады следующих друг за другом зон имеют разные знаки. Результирующий эффект в точке P получается суммированием вкладов от всех зон, т. е.

$$U(P) = 2i\lambda\, \frac{A \exp[ik(r_0 + b)]}{r_0 + b} \sum_{j=1}^{n} (-1)^{j+1} K_j. \tag{4}$$

Ряд

$$\sum = \sum_{j=1}^{n} (-1)^{j+1} K_j = K_1 - K_2 + K_3 - \ldots + (-1)^{n+1} K_n \tag{5}$$

можно приближенно вычислить по методу Шустера [2].

Перепишем вначале (5) в виде

$$\sum = \frac{K_1}{2} + \left(\frac{K_1}{2} - K_2 + \frac{K_3}{2}\right) + \left(\frac{K_3}{2} - K_4 + \frac{K_5}{2}\right) + \ldots \tag{6}$$

Последний член равен $^1/_2\, K_n$ или $^1/_2\, K_{n-1} - K_n$ в зависимости от того, нечетно n или четно. Допустим теперь, что закон, определяющий изменение K с направлением таков, что величина K_j *больше* среднего арифметического соседних значений K_{j-1} и K_{j+1}. Тогда каждый член в (6), заключенный в скобки, отрицателен и, значит,

$$\left.\begin{aligned} &\sum < \frac{K_1}{2} + \frac{K_n}{2}, && \text{если } n \text{ нечетно,} \\ &\sum < \frac{K_1}{2} + \frac{K_{n-1}}{2} - K_n, && \text{если } n \text{ четно.} \end{aligned}\right\} \tag{7}$$

и

Теперь перепишем (5) в виде

$$\sum = K_1 - \frac{K_2}{2} - \left(\frac{K_2}{2} - K_3 + \frac{K_4}{2}\right) - \left(\frac{K_4}{2} - K_5 + \frac{K_6}{2}\right) - \ldots \tag{8}$$

Последний член этого ряда равен $-\frac{1}{2} K_{n-1} + K_n$, если n нечетно, и $-\frac{1}{2} K_n$,

если оно четно. Следовательно,

$$\left.\begin{aligned} \sum > K_1 - \frac{K_2}{2} - \frac{K_{n-1}}{2} + K_n, &\quad \text{если } n \text{ нечетно,} \\ \text{и} \quad \sum > K_1 - \frac{K_2}{2} - \frac{K_n}{2}, &\quad \text{если } n \text{ четно.} \end{aligned}\right\} \quad (9)$$

Величина каждого K_j лишь немного отличается от величин соседних K_{j-1} и K_{j+1}, и поэтому правые части соотношений (7) и (9) практически равны и приближенно можно считать, что

$$\left.\begin{aligned} \sum = \frac{K_1}{2} + \frac{K_n}{2}, &\quad \text{если } n \text{ нечетно,} \\ \text{и} \quad \sum = \frac{K_1}{2} - \frac{K_n}{2}, &\quad \text{если } n \text{ четно.} \end{aligned}\right\} \quad (10)$$

Легко проверить, что соотношения (10) остаются справедливыми, если каждое K_j *меньше* среднего арифметического соседних членов, и тогда члены в (6) и (8), заключенные в скобках, отрицательны. Более того, можно ожидать, что (10) останется справедливым даже тогда, когда только часть членов в скобках отрицательна, а другая часть положительна. В этом случае ряд можно разделить на две части в зависимости от знаков членов в скобках, и к каждой такой части применить предыдущие рассуждения. Итак, мы приходим к заключению, что сумма ряда определяется выражениями (10), если члены в скобках в соотношениях (6) и (8) не меняют знак так часто, что неточности, складываясь, достигают значительной величины. Если исключить последний случай, то из (4) и (10) находим

$$U(P) = i\lambda (K_1 \pm K_n) \frac{A \exp[ik(r_0+b)]}{r_0+b}; \quad (11)$$

здесь верхний знак берется при нечетном n, а нижний — при четном.

Воспользовавшись (3), уравнение (11) можно переписать в виде

$$U(P) = \frac{1}{2}[U_1(P) + U_n(P)]. \quad (12)$$

Для последней зоны Z_n, видимой из P, QP становится касательной к волновому фронту, т. е. $\chi = \pi/2$, и, как говорилось выше, для такого χ величина K, по предположению, равна нулю. Следовательно, $K_n = 0$, и (11) сводится к выражению

$$U(P) = i\lambda K_1 \frac{A \exp[ik(r_0+b)]}{r_0+b} = \frac{1}{2} U_1(P), \quad (13)$$

показывающему, что *полное возмущение в P равняется половине возмущения, обусловленного действием первой зоны.*

Соотношение (13) находится в согласии с выражением, описывающим действие сферической волны, если

$$i\lambda K_1 = 1,$$

т. е. если

$$K_1 = -\frac{i}{\lambda} = \frac{\exp(-i\pi/2)}{\lambda}. \quad (14)$$

Множитель $\exp(-i\pi/2)$ можно объяснить, если предположить, что вторичные волны отстают по фазе на четверть периода от первичной волны. Присутствие другого множителя становится понятным, если допустить, что амплитуды вторичных и первичных волн относятся, как $1 : \lambda$. Таким образом, мы приходим к заключению, что при этих допущениях относительно амплитуды и фазы вторичных волн принцип Гюйгенса — Френеля правильно описывает распространение сферических волн в свободном пространстве. Однако приведенные выше дополнительные предположения нужно рассматривать просто как удобный способ интерпретации математических выражений; иными словами, они

не имеют какого-либо физического смысла. Истолкование множителя (14) станет очевидным позднее (см. § 8.3).

Следуя и дальше Френелю, закроем несколько зон плоским экраном, перпендикулярным к P_0P, с круглым отверстием, центр которого находится на этой линии, и рассмотрим действие оставшихся зон в точке P. Теперь следует считать, что суммарное возмущение в P обусловлено только волнами от незакрытых зон. Если экран оставляет открытой только половину первой зоны, то из (3), полагая $j=1$ и умножая на 1/2, получим

$$U(P)=i\lambda K_1\frac{A\exp[ik(r_0+b)]}{r_0+b}=\frac{A\exp[ik(r_0+b)]}{r_0+b}. \quad (15)$$

Следовательно, возмущение в P оказывается таким же, как и в отсутствие экрана. Если закрыты все зоны, кроме первой, то из (3) находим

$$U(P)=2i\lambda K_1\frac{A\exp[ik(r_0+b)]}{r_0+b}=2\frac{A\exp[ik(r_0+b)]}{r_0+b}, \quad (16)$$

и интенсивность $I(P)=|U(P)|^2$ в четыре раза больше, чем в отсутствие экрана. При дальнейшем увеличении отверстия интенсивность уменьшится, так как первые два члена в (4) имеют разные знаки. Больше того, K_1 и K_2 почти одинаковы, и, следовательно, если отверстие приблизительно равно двум первым зонам, то в точке P будет почти полная темнота. Поэтому при изменении размеров отверстия наблюдается периодическое изменение интенсивности в P. Такой же результат получается и тогда, когда размеры отверстия и источника остаются постоянными, а точка наблюдения P перемещается вдоль оси. В этом случае при постепенном приближении точки P к экрану увеличивается число открывающихся зон.

Все полученные выше результаты находятся в хорошем согласии с экспериментом. Одно из предсказаний теории Френеля произвело сильное впечатление на его современников и фактически положило конец долгому спору между сторонниками корпускулярной и волновой теорий света. Этот спор был решен в пользу волновой теории. Речь идет о явлении, наблюдаемом при закрытии первой зоны маленьким круглым диском, помещенным под прямым углом к прямой P_0P. Согласно (5) комплексная амплитуда в P равняется

$$U(P)=2i\lambda\frac{A\exp[ik(r_0+b)]}{r_0+b}[-K_2+K_3-K_4+\ldots]. \quad (17)$$

Так же, как и раньше, получим, что сумма ряда в скобках равна $-K_2/2$. Выше мы предположили, что K_2 только немного отличается от $K_1=1/i\lambda$ и, следовательно, в геометрической тени диска интенсивность света будет такой же, как и в отсутствие диска *).

§ 8.3. Теория дифракции Кирхгофа

8.3.1. Интегральная теорема Кирхгофа. Основная идея теории Гюйгенса—Френеля заключается в том, что световое возмущение в точке P возникает вследствие суперпозиции вторичных волн, испускаемых поверхностью, находящейся между этой точкой и источником света. Кирхгоф [3] придал этой

*) В 1818 г. Пуассон, будучи членом комитета Французской академии, рассматривавшего представленный на премию мемуар Френеля, показал на основании теории Френеля, что в центре тени маленького диска должно находиться светлое пятно. Этот результат Пуассон счел противоречащим опыту и тем самым отвергающим теорию Френеля. Однако Араго, другой член того же комитета, выполнил эксперимент, показавший, что это удивительное предсказание правильно. Такое же наблюдение, сделанное Маральди столетием раньше, было забыто.

идее строгий математический вид *) и показал, что принцип Гюйгенса — Френеля можно считать приближенной формой определенной интегральной теоремы **). В этой теореме решение однородного волнового уравнения в произвольной точке поля выражается через значения искомой величины и ее первой производной во всех точках произвольной замкнутой поверхности, окружающей точку P.

Рассмотрим сначала строго монохроматическую скалярную волну

$$V(x,\ y,\ z,\ t) = U(x,\ y,\ z)\exp(-i\omega t). \tag{1}$$

В вакууме ее часть, зависящая от координат, удовлетворяет волновому уравнению, не зависящему от времени,

$$(\nabla^2 + k^2)\,U = 0, \tag{2}$$

где $k = \omega/c$. Уравнение (2) называется также уравнением Гельмгольца.

Пусть v — объем, ограниченный произвольной замкнутой поверхностью S, а P — какая-нибудь точка внутри него; предположим, что U имеет непрерывные частные производные первого и второго порядков внутри этого объема и на поверхности S. Если U' — любая другая функция, удовлетворяющая таким же требованиям непрерывности, как и U, то по теореме Грина получим

$$\iiint\limits_{v} (U\nabla^2 U' - U'\nabla^2 U)\,dv = -\iint\limits_{S} \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) dS, \tag{3}$$

где $\partial/\partial n$ означает дифференцирование вдоль внутренней ***) нормали к поверхности S. В частности, если U' удовлетворяет также волновому уравнению, не зависящему от времени, т. е. если

$$(\nabla^2 + k^2)\,U' = 0, \tag{4}$$

то из (2) и (4) сразу же следует, что подынтегральное выражение в левой части (3) обращается в нуль в каждой точке объема v и, следовательно,

$$\iint\limits_{S} \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) dS = 0 \tag{5}$$

Рассмотрим функцию $U'(x,\ y,\ z) = e^{iks}/s$, где s — расстояние от P до точки $(x,\ y,\ z)$. Эта функция имеет особенность при $s = 0$, и так как предполагается, что U' непрерывна и дифференцируема, то, следовательно, точку P нужно исключить из области интегрирования. Поэтому окружим ее небольшой сферой радиуса ε и произведем интегрирование по объему, заключенному между S и поверхностью этой сферы S' (рис. 8.2). Тогда вместо (5) получим

$$\iint\limits_{S} + \iint\limits_{S'} \left\{U\frac{\partial}{\partial n}\left(\frac{\exp(iks)}{s}\right) - \frac{\exp(iks)}{s}\frac{\partial U}{\partial n}\right\} dS = 0,$$

*) Теория Кирхгофа применима к дифракции скалярных волн. В § 8.4 будет показано, что скалярная теория обычно вполне пригодна при рассмотрении проблем инструментальной оптики.

Векторное обобщение принципа Гюйгенса — Френеля предлагалось многими авторами. Первая удовлетворительная работа, посвященная этому вопросу, принадлежит Коттлеру [4] (см. [5], стр. 144).

**) Для монохроматических волн эта теорема была выведена раньше в акустике Гельмгольцем [6].

***) Теорема Грина обычно формулируется для внешней нормали, но в данном случае более удобна внутренняя нормаль.

откуда

$$\iint\limits_S \left\{U \frac{\partial}{\partial n}\left(\frac{\exp(iks)}{s}\right) - \frac{\exp(iks)}{s}\frac{\partial U}{\partial n}\right\} dS = -$$

$$= -\iint\limits_{S'} \left\{U \frac{\exp(iks)}{s}\left(ik - \frac{1}{s}\right) - \frac{\exp(iks)}{s}\frac{\partial U}{\partial n}\right\} dS' =$$

$$= -\iint\limits_{\Omega} \left\{U \frac{\exp(ik\varepsilon)}{\varepsilon}\left(ik - \frac{1}{\varepsilon}\right) - \frac{\exp(ik\varepsilon)}{\varepsilon}\frac{\partial U}{\partial s}\right\} \varepsilon^2 d\Omega, \quad (6)$$

где $d\Omega$ — элемент телесного угла. Так как интеграл по S не зависит от ε, можно заменить интеграл в правой части (6) его предельным значением при $\varepsilon \to 0$. Первый и третий члены в нем не дают вклада в этот предел, а полный вклад второго члена равен $4\pi U(P)$. Следовательно,

$$U(P) = \frac{1}{4\pi}\iint\limits_S \left\{U \frac{\partial}{\partial n}\left(\frac{\exp(iks)}{s}\right) - \frac{\exp(iks)}{s}\frac{\partial U}{\partial n}\right\} dS. \quad (7)$$

Рис. 8.2. К выводу интегральной теоремы Гельмгольца — Кирхгофа; область интегрирования.

Это одна из форм *интегральной теоремы Гельмгольца и Кирхгофа* *).

Заметим, что когда $k \to 0$, то не зависящее от времени волновое уравнение (3) сводится к уравнению Лапласа $\nabla^2 U = 0$, и (7) переходит тогда в хорошо известную формулу теории потенциала

$$U(P) = \frac{1}{4\pi}\iint\limits_S \left\{U \frac{\partial}{\partial n}\left(\frac{1}{s}\right) - \frac{1}{s}\frac{\partial U}{\partial n}\right\} dS. \quad (8)$$

Если P лежит вне поверхности S, но U — по-прежнему непрерывная и дифференцируемая до второго порядка функция внутри S и если, как и раньше, принять $U' = \exp(iks)/s$, то уравнение (3) остается справедливым по всему объему внутри S. Тогда, согласно (5), интеграл по поверхности равен нулю.

Существует другая дополнительная форма теоремы Гельмгольца — Кирхгофа для случая, когда функция U непрерывна и дифференцируема до второго порядка *вне* и на самой замкнутой поверхности S (источники внутри). Однако в таком случае, как и в задачах, связанных с распространением света в бесконечной среде, одних граничных значений на S уже недостаточно для получения однозначного решения. Здесь требуются еще дополнительные предположения относительно решения **) при $s \to \infty$.

До сих пор рассматривались только строго монохроматические волны. Теперь выведем теорему Кирхгофа в общем виде, пригодном и в случае немонохроматических волн.

Пусть $V(x, y, z, t)$ — решение волнового уравнения

$$\nabla^2 V = \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2}, \quad (9)$$

*) Эта теорема выражает $U(P)$ через U и $\partial U/\partial n$ на S. Тем не менее можно показать на основании теории функций Грина, что для определения U в каждой точке P внутри S достаточно одной какой-нибудь величины, т. е. или U, или $\partial U/\partial n$ (см., например, [7]). Однако только в простейших случаях, например когда S — плоскость, можно найти соответствующую функцию Грина (см. [8]).

**) Желающие более подробно познакомиться с этим вопросом могут воспользоваться, например, [5].

и его можно представить в виде интеграла Фурье

$$V(x,\ y,\ z,\ t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} U_\omega(x,\ y,\ z)\exp(-i\omega t)\,d\omega. \qquad (10)$$

Тогда по формуле обратного фурье-преобразования

$$U_\omega(x,\ y,\ z)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} V(x,\ y,\ z,\ t)\exp(i\omega t)\,dt. \qquad (11)$$

Так как предполагается, что $V(x,\ y,\ z,\ t)$ удовлетворяет волновому уравнению (9), то $U_\omega(x,\ y,\ z)$ удовлетворяет не зависящему от времени волновому уравнению (2). Если, кроме того, V подчиняется соответствующим условиям регулярности внутри замкнутой поверхности S и на ней, то мы вправе применить формулу Кирхгофа отдельно к каждой фурье-компоненте $U_\omega(x,\ y,\ z)=U_\omega(P)$, т. е. написать

$$U_\omega(P)=\frac{1}{4\pi}\iint\limits_S\left\{U_\omega\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial U_\omega}{\partial n}\right\}dS. \qquad (12)$$

Изменяя порядок интегрирования и полагая $k=\omega/c$, приведем (10) к виду

$$V(P,\ t)=$$

$$=\frac{1}{4\pi}\iint\limits_S dS\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}\left\{U_\omega\frac{\partial}{\partial n}\left(\frac{\exp[-i\omega(t-s/c)]}{s}\right)-\frac{\exp[-i\omega(t-s/c)]}{s}\frac{\partial U_\omega}{\partial n}\right\}d\omega=$$

$$=\frac{1}{4\pi}\iint\limits_S dS\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}\left\{U_\omega\left\{\frac{\partial}{\partial n}\left(\frac{1}{s}\right)+\frac{i\omega}{sc}\frac{\partial s}{\partial n}\right\}\exp[-i\omega(t-s/c)]-\right.$$

$$\left.-\frac{\exp[-i\omega(t-s/c)]}{s}\frac{\partial U_\omega}{\partial n}\right\}d\omega,$$

или, используя (10)

$$V(P,\ t)=\frac{1}{4\pi}\iint\limits_S\left\{[V]\frac{\partial}{\partial n}\left(\frac{1}{s}\right)-\frac{1}{cs}\frac{\partial s}{\partial n}\left[\frac{\partial V}{\partial t}\right]-\frac{1}{s}\left[\frac{\partial V}{\partial n}\right]\right\}dS. \qquad (13)$$

В квадратных скобках заключены «запаздывающие величины», т. е. значения функций, взятых в момент времени $t-s/c$. Формула (13) представляет собой *теорему Кирхгофа* в общем виде.

По аналогии с предыдущим случаем отметим, что если P находится вне S, то величина интеграла в (13) равна нулю.

Последний член в (13) представляет собой вклад в решение, обусловленный распределением источников с «силой» $-\frac{1}{4\pi}\frac{\partial V}{\partial n}$ на единицу площади, а первые два члена — вклад от диполей «силой» $V/4\pi$ на единицу площади, направленных нормально к поверхности. Естественно, что эти источники и диполи фиктивные и, следовательно, в такой интерпретации нет глубокого физического смысла.

8.3.2. Теория дифракции Кирхгофа. Интегральная теорема Кирхгофа базируется на основной идее принципа Гюйгенса — Френеля. Однако законы, управляющие вкладами от различных элементов поверхности, значительно сложнее, чем предполагал Френель. Тем не менее Кирхгоф показал, что во многих случаях эту теорему можно свести к приближенной, но более простой форме, эквивалентной формулировке Френеля, и, кроме того, определить точный вид коэффициента наклона, который в теории Френеля остается неопределенным.

Рассмотрим монохроматическую волну, идущую от точечного источника P_0 сквозь отверстие в плоском непрозрачном экране. Пусть, как и раньше, P — точка, в которой определяется световое возмущение. Допустим, что линейные размеры отверстия велики по сравнению с длиной волны света, но малы по сравнению с расстояниями от P_0 и P до экрана.

Для того чтобы найти возмущение в точке P, рассмотрим интеграл Кирхгофа по поверхности S, образованной (рис. 8.3, *а*): 1) отверстием $\mathcal{A}$, 2) участком $\mathcal{B}$ неосвещенной стороны экрана и 3) частью $\mathcal{C}$ большой сферы с центром

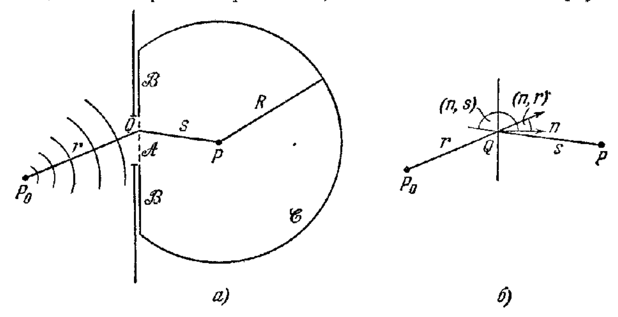


Рис. 8.3. К выводу дифракционной формулы Френеля — Кирхгофа.

в P и радиусом R, которая вместе с $\mathcal{A}$ и $\mathcal{B}$ образует замкнутую поверхность.

Из теоремы Кирхгофа в форме (7) имеем

$$U(P)=\frac{1}{4\pi}\left[\iint\limits_{\mathcal{A}}+\iint\limits_{\mathcal{B}}+\iint\limits_{\mathcal{C}}\right]\left\{U\frac{\partial}{\partial n}\left(\frac{\exp(iks)}{s}\right)-\left(\frac{\exp(iks)}{s}\right)\frac{\partial U}{\partial n}\right\}dS, \tag{14}$$

где, как и раньше, s — расстояние между элементом dS и P, а $\partial/\partial n$ — обозначает дифференцирование вдоль внутренней нормали к поверхности интегрирования.

Здесь дело осложняется тем, что значения U и $\partial U/\partial n$ на $\mathcal{A}$, $\mathcal{B}$ и $\mathcal{C}$, которые необходимо подставить в (14), никогда точно неизвестны. Однако разумно считать, что повсюду на $\mathcal{A}$, кроме мест, находящихся в непосредственной близости к краю отверстия, U и $\partial U/\partial n$ мало отличаются от тех значений, которые они имели бы в отсутствие экрана, и что на $\mathcal{B}$ эти величины близки к нулю. Тогда, согласно Кирхгофу, имеем

$$\left.\begin{aligned}&\text{на } \mathcal{A}:\ U=U^{(i)}, && \frac{\partial U}{\partial n}=\frac{\partial U^{(i)}}{\partial n},\\ &\text{на } \mathcal{B}:\ U=0, && \frac{\partial U}{\partial n}=0,\end{aligned}\right\} \tag{15}$$

где

$$U^{(i)}=\frac{Ae^{ikr}}{r},\qquad \frac{\partial U^{(i)}}{\partial n}=\frac{Ae^{ikr}}{r}\left[ik-\frac{1}{r}\right]\cos(n, r) \tag{16}$$

— величины, относящиеся к падающему полю (рис. 8.3, *б*), а A — постоянная Приближения (15), называемые *граничными условиями Кирхгофа*, лежат в основе *теории дифракции Кирхгофа*

Остается еще учесть роль части сферической поверхности $\mathcal{C}$. Теперь очевидно, что, беря радиус R достаточно большим, можно получить значения U и $\partial U/\partial n$ на $\mathcal{C}$ сколь угодно малыми и, следовательно, можно пренебречь вкладом от $\mathcal{C}$. Однако при неограниченном увеличении R площадь $\mathcal{C}$ также неограниченно увеличивается и условие $U\to 0$ и $\partial U/\partial n\to 0$ при $R\to\infty$ недостаточно для того, чтобы наш интеграл стремился к нулю. Таким образом, необходимы

более точные допущения относительно поведения волновой функции на большом расстоянии от экрана. Этот вопрос уже обсуждался на стр. 347 в связи с вопросом об однозначности решений задач, рассматривающих бесконечную среду. Для нашей же задачи достаточно сделать физически очевидное допущение, что радиационное поле не существовало всегда, а начало создаваться источником в некоторый определенный момент времени *) $t = t_0$. (Это, конечно, означает отступление от строгой монохроматичности, так как идеально монохроматическое поле существует неограниченное время.) Тогда в любой момент времени $t > t_0$ поле заполняет некоторую часть пространства, внешняя граница которой находится от P_0 на расстоянии, не превышающем $c(t - t_0)$, где c — скорость света. Следовательно, если радиус R выбирается столь большим, что в момент наблюдения в P вклад в возмущение от $\mathscr{C}$ отсутствует (так как в этот момент поле еще не достигло столь удаленных областей), то интеграл по $\mathscr{C}$ равен нулю. Учитывая это и пренебрегая в производных по нормали членами $1/r$ и $1/s$, малыми по сравнению с k, окончательно получим вместо (14)

$$U(P) = -\frac{iA}{2\lambda} \iint\limits_{\mathscr{A}} \frac{\exp[ik(r+s)]}{rs} [\cos(n, r) - \cos(n, s)]\, dS. \tag{17}$$

Это выражение называется *дифракционной формулой Френеля — Кирхгофа*.

Очевидно, что вместо $\mathscr{A}$ мы вправе выбрать любую другую незамкнутую поверхность, границы которой совпадают с краем отверстия. В частности, вместо $\mathscr{A}$ можно взять часть W падающего волнового фронта, которая приблизительно заполняет отверстие, и часть $\mathscr{C}$ конуса с вершиной в P_0 и с образующими, проходящими через края отверстия (рис. 8.4). При достаточно большом радиусе кривизны волнового фронта вкладом от $\mathscr{C}$, очевидно, можно пренебречь. Кроме того, на W имеем $\cos(n, r_0) = 1$. Если еще положить $\chi = \pi - (r_0, s)$, то вместо (17) получим

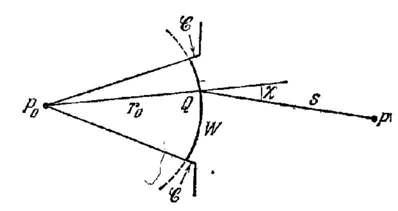


Рис. 8.4. К выводу дифракционной формулы (18).

$$U(P) = -\frac{i}{2\lambda} \frac{A\exp(ikr_0)}{r_0} \iint\limits_{W} \frac{\exp(iks)}{s} (1 + \cos\chi)\, dS, \tag{18}$$

где r_0 — радиус волнового фронта W. Этот результат находится в согласии с формулировкой принципа Гюйгенса Френелем, если вкладом от элемента dW волнового фронта считать

$$-\frac{i}{2\lambda} \frac{A\exp(ikr_0)}{r_0} \frac{\exp(iks)}{s} (1 + \cos\chi) dS. \tag{19}$$

Сравнивая (18) с (8.2.1), найдем для коэффициента наклона, фигурирующего в теории Френеля, выражение **)

$$K(\chi) = -\frac{i}{2\lambda} (1 + \cos\chi). \tag{20}$$

Для центральной зоны $\chi = 0$ и (20) дает $K_1 = K(0) = -i/\lambda$, что согласуется с (8.2.14). Видно, однако, что Френель неправильно предполагал, будто $K(\pi/2) = 0$.

Возвращаясь снова к дифракционной формуле Френеля — Кирхгофа (17), отметим, что она симметрична относительно источника и точки наблюдения. Это

*) Это предположение несущественно, но сокращает обсуждение. Более формальную аргументацию см. в [9].

**) Выражение (20) для коэффициента наклона впервые получено Стоксом [10].

означает, что *точечный источник, находящийся в P_0, производит в P такое же действие, какое производил бы точечный источник равной интенсивности, помещенной в P.* Этот вывод иногда называют *теоремой взаимности* (или *теоремой обратимости*) *Гельмгольца.*

До сих пор мы предполагали, что свет на пути от источника до точки P не встречает других поверхностей, кроме дифракционного экрана: в таком случае падающие волны сферические. Легко распространить этот анализ и на более сложные случая, когда форма волны не столь проста. И тогда мы опять получим, что выводы теории Кирхгофа по существу эквивалентны предсказаниям, сделанным на основе принципа Гюйгенса — Френеля, при условии, что в каждой точке волнового фронта радиусы его кривизны велики по сравнению с длиной волны света, а углы достаточно малы.

Из предыдущих рассуждений можно сразу же вывести заключение о распределении света, дифрагировавшего на дополнительных друг другу экранах, т. е. на экранах, у которых отверстия одного точно совпадают с непрозрачными частями другого и наоборот. Пусть $U_1(P)$ и $U_2(P)$ — комплексные возмущения, когда только один из экранов помещен на пути между источником и точкой наблюдения P. Тогда, поскольку U_1 и U_2 можно представить в виде интегралов по отверстиям, а отверстия в дополнительных экранах располагаются так, что полностью «открывают» весь волновой фронт, то

$$U_1+U_2=U. \tag{21}$$

Это так называемый *принцип Бабине* *) [11].

Из принципа Бабине можно вывести два заключения. Если $U_1=0$, то $U_2=U$, т. е. в точках, где интенсивность при наличии одного экрана равна нулю, в присутствии лишь другого экрана она будет такой же, как и в отсутствие экранов. Далее, если $U=0$, то $U_1=-U_2$, т. е. в точках, где U равно нулю, фазы U_1 и U_2 различаются на π, а интенсивности $I_1=|U_1|^2$ и $I_2=|U_2|^2$ одинаковы. Так, например, если точечный источник изображается хорошо коррегированным объективом, распределение света U в плоскости изображений повсюду равно нулю, за исключением мест, находящихся в непосредственной близости от изображения O источника. Если дополнительные экраны поместить на пути между источником и изображением, то $I_1=I_2$ всюду, за исключением мест близ O.

Выводы из основного приближения (15) теории Кирхгофа подвергались многим критическим замечаниям, из которых следует, например, что решение Кирхгофа не дает исходных значений интенсивности в плоскости отверстия [12] (см. также [5] стр. 71, 72 и [13]).

Однако сравнительно недавно, Вольф и Марчанд [14] показали, что теорию Кирхгофа можно изложить полностью математически **). В таком виде теория дает точное решение некоторых иных краевых задач, чем (15) и (16), и полностью применима к основным проблемам инструментальной оптики. Это объясняется главным образом тем, что длины волн оптического диапазона малы по сравнению с размерами препятствий, на которых происходит дифракция [17] В других задачах, относящихся, например, к поведению поля в непосредственной близости к экранам и другим препятствиям, нужно применять более

*) В § 11.3 приводится аналогичная теорема, которую можно считать строгой формулировкой принципа Бабине. В этой теореме рассматривается не только скалярная функция U, но и векторное электромагнитное поле.

**) Коттлер [14а] также нашел, что теория Кирхгофа дает строгое решение определенной задачи, связанной с разрывом непрерывности (задачи с заданными разрывами, а не с заданными граничными значениями). Такая интерпретация представляет большой интерес в связи с задачей дифракции на черном (полностью поглощающем) экране (см. также [5], стр. 98 и [15]).

В статье [16] содержатся многочисленные ссылки на работы, касающиеся различных видоизменений теории Кирхгофа.

тонкие методы. Такие задачи необходимо рассматривать как задачи электромагнитной теории с граничными условиями и считать источники особыми точками волновых функций. Решения подобных задач найдены только для очень небольшого числа случаев; некоторые из них будут рассмотрены в гл 11.

8.3.3. Дифракция Фраунгофера и Френеля. Исследуем теперь подробно дифракционный интеграл Френеля — Кирхгофа (17)

$$U(P) = -\frac{Ai}{2\lambda}\iint_{\mathcal{A}} \frac{\exp[ik(r+s)]}{rs}[\cos(n, r) - \cos(n, s)]\,dS. \tag{22}$$

Когда элемент dS пробегает область интегрирования, $r+s$ в общем случае будет изменяться на очень много длин волн; и поэтому множитель в $\exp[ik(r+s)]$ будет быстро осциллировать. Кроме того, если расстояния от точек P_0 и P до экрана велики по сравнению с линейными размерами отверстия, то множитель $[\cos(n, r) - \cos(n, s)]$ изменяется по отверстию незначительно. Далее, предполагая, что O это любая точка отверстия и углы, образованные линиями P_0O и OP с P_0P, не слишком велики, можно заменить этот множитель на $2\cos\delta$, где δ — угол между линией P_0P и нормалью к экрану И наконец, множитель $1/rs$ можно заменить на $1/r's'$, где r' и s' — расстояния от точек P_0 и P до начала координат; тогда (22) сведется к

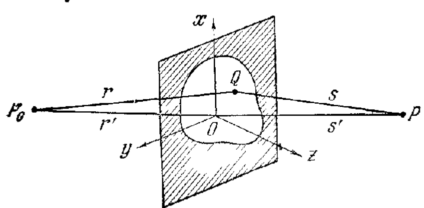


Рис. 8 5. Дифракция на отверстии в плоском экране

$$U(P) \approx -\frac{Ai}{\lambda}\frac{\cos\delta}{r's'}\iint_{\mathcal{A}} \exp[ik(r+s)]\,dS \tag{23}$$

Возьмем за начало декартовой системы координат точку O отверстия, а оси Ox и Oy выберем в плоскости отверстия. Будем считать, что ось z направлена в сторону полупространства, в котором находится точка наблюдения P (рис. 8.5).

Пусть (x_0, y_0, z_0) и (x, y, z) — координаты точек P_0 и P соответственно, а (ξ, η) — координаты точки Q отверстия; тогда

$$\left.\begin{aligned} r^2 &= (x_0-\xi)^2 + (y_0-\eta)^2 + z_0^2, \\ s^2 &= (x-\xi)^2 + (y-\eta)^2 + z^2, \end{aligned}\right\} \tag{24}$$

$$\left.\begin{aligned} r'^2 &= x_0^2 + y_0^2 + z_0^2, \\ s'^2 &= x^2 + y^2 + z^2. \end{aligned}\right\} \tag{25}$$

Следовательно,

$$\left.\begin{aligned} r^2 &= r'^2 - 2(x_0\xi + y_0\eta) + \xi^2 + \eta^2, \\ s^2 &= s'^2 - 2(x\xi + y\eta) + \xi^2 + \eta^2. \end{aligned}\right\} \tag{26}$$

Мы предположили, что линейные размеры экрана малы по сравнению с r' и s', и поэтому можем разложить r и s в степенные ряды по ξ/r', η/r', ξ/s' и η/s'; тогда получим

$$\left.\begin{aligned} r &\approx r' - \frac{x_0\xi + y_0\eta}{r'} + \frac{\xi^2+\eta^2}{2r'} - \frac{(x_0\xi + y_0\eta)^2}{2r'^3} - \ldots \\ s &\approx s' - \frac{x\xi + y\eta}{s'} + \frac{\xi^2+\eta^2}{2s'} - \frac{(x\xi + y\eta)^2}{2s'^3} - \ldots \end{aligned}\right\} \tag{27}$$

Подстановка (27) в (23) дает

$$U(P) = -\frac{i\cos\delta}{\lambda}\frac{A\exp[ik(r'+s')]}{r's'}\iint_{\mathcal{A}} \exp[ikf(\xi, \eta)]\,d\xi\,d\eta, \tag{28}$$

где

$$f(\xi,\ \eta)=-\frac{x_0\xi+y_0\eta}{r'}-\frac{x\xi+y\eta}{s'}+\frac{\xi^2+\eta^2}{2r}+\frac{\xi^2+\eta^2}{2s'}-\frac{(x_0\xi+y_0\eta)^2}{2r'^3}-\frac{(x\xi+y\eta)^2}{2s'^3}\ \dots \quad (29)$$

Если мы обозначим первые два направляющих косинуса через $(l_0,\ m_0)$ и $(l,\ m)$, т. е.

$$l_0=-\frac{x_0}{r'},\quad l=\frac{x}{s'},\quad m_0=-\frac{y_0}{r'},\quad m=\frac{y}{s'}, \quad (30)$$

то (29) можно переписать в виде

$$f(\xi,\ \eta)=(l_0-l)\,\xi+(m_0-m)\,\eta+\frac{1}{2}\left\{\left(\frac{1}{r'}+\frac{1}{s'}\right)(\xi^2+\eta^2)-\right.$$
$$\left.-\frac{(l_0\xi+m_0\eta)^2}{r'}-\frac{(l\xi+m\eta)^2}{s'}\right\}\dots \quad (31)$$

Мы свели задачу определения светового возмущения в P к вычислению интеграла (28). Конечно, оно упростится, если в f пренебречь квадратичными членами и членами более высоких порядков относительно ξ и η. В этом случае мы имеем дело с *дифракцией Фраунгофера*; если же квадратичными членами пренебречь нельзя,— то с *дифракцией Френеля*. К счастью, более простой случай дифракции Фраунгофера представляется в оптике значительно более важным.

Строго говоря, члены второго и более высоких порядков исчезают только в предельном случае $r'\to\infty$ и $s'\to\infty$, т. е. когда и источник, и точка наблюдения находятся в бесконечности (тогда надо допустить, что и множитель A перед интегралом стремится к бесконечности так же, как $r's'$. Однако очевидно, что вклад членов второго порядка в интеграл невелик, если

$$\frac{1}{2}k\left|\left(\frac{1}{r'}+\frac{1}{s'}\right)(\xi^2+\eta^2)-\frac{(l_0\xi+m_0\eta)^2}{r'}-\frac{(l\xi+m\eta)^2}{s'}\right|\ll 2\pi. \quad (32)$$

Можно сразу же указать определенные условия, при которых (32) удовлетворяется. Для этого воспользуемся неравенством вида $(l_0\xi+m_0\eta)^2\leqslant$ $\leqslant(l_0^2+m_0^2)(\xi^2+\eta^2)$ и, вспомнив, что l_0^2, m_0^2, l^2 и m^2 не превышают единицы, получим, что (32) удовлетворяется, если

$$|r'|\gg\frac{(\xi^2+\eta^2)_{\text{макс}}}{\lambda}\quad \text{и}\quad |s'|\gg\frac{(\xi^2+\eta^2)_{\text{макс}}}{\lambda}, \quad (33)$$

или если

$$\frac{1}{r'}+\frac{1}{s'}=0\quad \text{и}\quad l_0^2,\ m_0^2,\ l^2,\ m^2\ll\frac{|r'|\lambda}{(\xi^2+\eta^2)_{\text{макс}}}. \quad (34)$$

Условия (33) позволяют оценить расстояния r' и s', при которых применимо приближение Фраунгофера. Условия (34) означают, что дифракция Фраунгофера имеет место и тогда, когда точка наблюдения находится в плоскости, параллельной плоскости отверстия при условии, что точка наблюдения и источник света достаточно близки к оси z. Здесь следует различать два случая. Если r' отрицательно, то падающие на отверстие волновые фронты имеют вогнутость в направлении распространения и точка P_0 является центром схождения, а не расхождения падающей волны. Этот случай очень важен для практики, так как осуществляется в пространстве изображений хорошо корригированной центрированной системы, изображающей точечный источник, расположенный недалеко от оси. Дифракционная картина Фраунгофера образуется в параксиальной плоскости изображений и может рассматриваться как результат дифракции, дающей изображения волны на выходном зрачке. Если r' положительно, то волновые фронты имеют выпуклости в направлении распространения; дифракционные картины оказываются мнимыми и кажутся образованными на экране, проходящем через источник P_0. Этот случай имеет место, например, тогда, когда отверстие в экране находится непосредственно

перед глазом или когда установка объектива зрительной трубы соответствует рассматриванию удаленного источника света.

Чтобы составить ясное физическое представление о том, почему дифракционная картина Фраунгофера наблюдается в фокальной плоскости хорошо коррегированного объектива, сравним сначала две ситуации, показанные на рис. 8.6. На рис. 8.6, *а* пучок лучей от бесконечно удаленной точки падает на отверстие в направлении, определяемом направляющими косинусами l_0, m_0, n_0. Можно считать, что дифракция, наблюдаемая в направлении l, m, n в очень удаленной точке P, возникла в результате суперпозиции плоских волн, исходящих из каждой точки отверстия в этом направлении. Такие волны (не

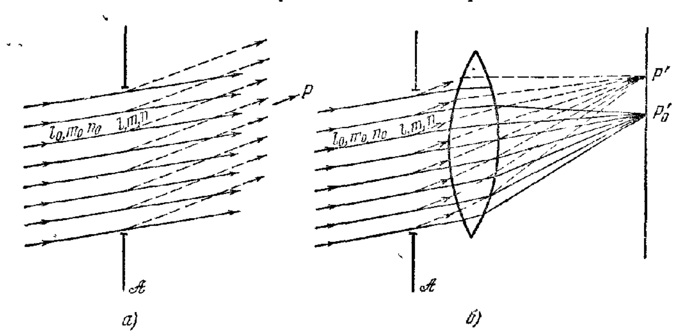


Рис. 8.6. Сравнение двух случаев дифракции Фраунгофера.

существующие в рамках геометрической оптики) можно назвать *дифрагировавшими волнами*, а соответствующие волновые нормали — *дифрагировавшими лучами*.

Если теперь поместить хорошо коррегированный объектив позади экрана (см. рис. 8.6, *б*), то весь свет, дифрагировавший в направлении (l, m, n), соберется в фокусе P' в фокальной плоскости объектива. Так как длины оптических путей всех лучей, приходящих в P' от волнового фронта дифрагировавшего пучка, равны, то по существу интерференционные эффекты остаются такими же, как и в первом случае, конечно, при условии, что объектив так велик, что не вносит дополнительной дифракции. Более общее ограничение, состоящее в том, что на отверстие должна падать плоская волна, также можно снять, если длины путей от источника до P' примерно одинаковы для всех лучей.

В случае дифракции Фраунгофера четыре величины l_0, m_0, l, m входят в (31) только в комбинации

$$p = l - l_0, \qquad q = m - m_0. \tag{35}$$

Следовательно, в той области, где справедливо упомянутое выше приближение, картина не изменится, если отверстие сместится в своей собственной плоскости.

Запишем интеграл, описывающий дифракцию Фраунгофера, в виде

$$U(P) = C \iint_{\mathcal{A}} \exp\left[-ik(p\xi + q\eta)\right] d\xi\, d\eta; \tag{36}$$

здесь C — величина, стоящая перед интегралом (28). C определяется через величины, связанные с положениями источника и точки наблюдения; однако на практике часто удобнее выражать ее через другие величины. Пусть E — полная энергия, падающая на отверстие. По закону сохранения энергии вся энергия, достигающая плоскости наблюдения, должна равняться E; поэтому

должно выполняться нормирующее условие

$$\iint \left| U(p,\ q) \right|^2 dp\, dq = E; \tag{37}$$

здесь интегрирование производится по всем возможным значениям величин p и q. Уравнение (36) можно записать теперь в виде интеграла Фурье

$$U(p,\ q) = \iint G(\xi,\ \eta) \exp\left[-\frac{2\pi i}{\lambda}(p\xi + q\eta)\right] d\xi\, d\eta, \tag{38}$$

где G — *функция зрачка* *), определяемая как

$$\left.\begin{aligned} &G(\xi,\ \eta) = \mathrm{const}\,(C) \ \text{в точках отверстия,} \\ &G(\xi,\ \eta) = 0 \qquad\qquad \text{в точках вне отверстия,} \end{aligned}\right\} \tag{39}$$

а интеграл берется по всей $(\xi,\ \eta)$-плоскости.

По теореме Парсеваля для фурье-преобразования [18] имеем

$$\iint |G(\xi,\ \eta)|^2 d\xi\, d\eta = \left(\frac{1}{\lambda}\right)^2 \iint |U(p,\ q)|^2 dp\, dq, \tag{40}$$

или, используя (37) и (39) и обозначая площадь отверстия через D,

$$\frac{1}{\lambda^2} E = |C|^2 D, \tag{41}$$

откуда **)

$$C = \frac{1}{\lambda} \sqrt{\frac{E}{D}}. \tag{42}$$

Тогда, основной интеграл, описывающий дифракцию Фраунгофера, принимает вид

$$U(p,\ q) = \frac{1}{\lambda} \sqrt{\frac{E}{D}} \iint \exp\left[-ik(p\xi + q\eta)\right] d\xi\, d\eta. \tag{43}$$

Следует обратить внимание, что интенсивность $I_0 = |U(0,\ 0)|^2$ в центре картины, где $p = q = 0$, равна

$$I_0 = \left(\frac{1}{\lambda}\right)^2 \frac{E}{D} \left(\iint d\xi\, d\eta\right)^2 = \frac{ED}{\lambda^2} = C^2 D^2. \tag{44}$$

При выводе (43) мы пренебрегли тем, что формула (36) была получена для ограниченной области величин p и q. Однако ошибка, вносимая при распространении интегрирования в (40) на все значения p и q, ничтожна, так как величина $U(p,\ q)$ очень мала всюду, за исключением области, находящейся в непосредственной близости от точки $p = q = 0$.

Однако вернемся к основному интегралу (28) теории дифракции. Когда точка $(\xi,\ \eta)$ пробегает область интегрирования, функции $f(\xi,\ \eta)$ изменяется на очень много длин волн; поэтому вещественная и мнимая части подынтегрального выражения многократно изменяют знак. В общем случае вклады от различных элементов фактически уничтожают друг друга (деструктивная интерференция). Но для элементарного участка, окружающего точку (назовем ее *критической точкой* или *полюсом*), где $f(\xi,\ \eta)$ постоянна, положение другое. Здесь подынтегральное выражение изменяется значительно медленнее, и можно ожидать, что его вклад станет заметным. Поэтому, если длина волны достаточно мала, величина интеграла, по существу, определяется поведением f вблизи точек, где f постоянно. Это является основой *метода стационарной фазы*, позволяющего определить асимптотическое поведение интегралов определенного класса (более подробно он разбирается в приложении 3). Ниже мы

*) Более общие функции зрачка рассматриваются в § 8.6 и 9.5.

**) Постоянный фазовый множитель опускается, так как он не вносит вклада в интенсивность $I = |U|^2$.

используем предыдущие результаты для классификации дифракционных явлений.

Сравнивая (22) и (28), мы видим, что $r+s=r'+s'+f$, и поэтому (см. рис. 8.5)

$$f = P_0Q + QP + \text{const}. \tag{45}$$

Очевидно, f, рассматриваемая как функция Q, постоянна, когда Q коллинеарно с P_0 и P. Следовательно, основной вклад в возмущение в точке P поступает от точек, находящихся в непосредственной близости к точке $\bar{Q}$, в которой линия, соединяющая источник с точкой наблюдения, пересекает плоскость отверстия. В частном случае дифракции Фраунгофера, P_0 и P находятся в бесконечности и особой точки $\bar{Q}$ не существует. Следовательно, в данном случае поведение дифракционного интеграла должно быть необычным.

В § 8.5—8.8 мы рассмотрим наиболее важные случаи дифракции Фраунгофера и Френеля. Но сначала следует убедиться, что при вычислении интенсивности света мы вправе пользоваться скалярной волновой функцией U.

§ 8.4. Переход к скалярной теории*)

При выводе интегральной теоремы Кирхгофа мы воспользовались только одним свойством функции U, а именно тем, что она удовлетворяет однородному скалярному волновому уравнению. Следовательно, эта теорема и заключения предыдущей главы применимы к каждой декартовой компоненте векторов поля, векторного потенциала, векторов Герца и т. д. в областях, где не существует ни токов, ни зарядов. Для того чтобы полностью описать поле, теорему Кирхгофа следует применять отдельно к каждой декартовой компоненте. Однако в силу удачного стечения обстоятельств в большинстве оптических задач вполне достаточно приближенного описания поля одной комплексной скалярной волновой функцией.

Для полного описания электромагнитного поля необходимо знать величины векторов поля, а также их направления (поляризацию) как функции положения и времени. Но в оптическом поле вследствие его очень высокой частоты (порядка 10^{14} *сек*$^{-1}$) измеряются не мгновенные значения этих величин, а лишь значения, усредненные по времени в интервале, большом по сравнению с периодом световых колебаний. Более того, обычно имеют дело с естественным светом, где в наблюдаемом (макроскопическом) поле нет преимущественных направлений поляризации. В этом случае первостепенное значение имеет *интенсивность* I, уже определенная в п. 1.1.4 как средняя по времени энергия, протекающая в единицу времени через единицу площади, содержащую электрический и магнитный векторы, т. е.

$$I = \frac{c}{4\pi} | \langle \mathbf{E} \times \mathbf{H} \rangle |.$$

Мы покажем, что интенсивность электромагнитного поля, связанного с прохождением естественного света через правильно рассчитанные оптические приборы с отверстием средней величины, можно приближенно выразить через одну комплексную скалярную волновую функцию с помощью формулы **)

$$I = |U|^2$$

и что функцию U можно вычислить, зная эйконал системы.

8.4.1. Поле изображения, создаваемое монохроматическим осциллятором. Рассмотрим симметричную оптическую систему, освещенную точечным источ-

*) Здесь мы в основном пользуемся исследованием [19]

**) Вольф [20] показал, что в более общем случае усредненные по времени плотность энергии и поток энергии в поле неполяризованного квазимонохроматического излучения всегда можно получить из одной комплексной гармонической во времени скалярной волновой функции.

ником света, находящимся в P_0 (рис. 8.7) и испускающим естественный квазимонохроматический свет с частотой ω_0. Предположим, что угол наклона лучей, прошедших сквозь систему, к оси невелик, скажем, не больше 10° или 15°. Поместим начало декартовой системы координат (x_1, x_2, x_3) в точку P_0 и будем считать, что ось x_3 направлена вдоль главного луча. Можно рассматривать источник как диполь с моментом $\mathbf{Q}(t)$, у которого со временем t изменяется и

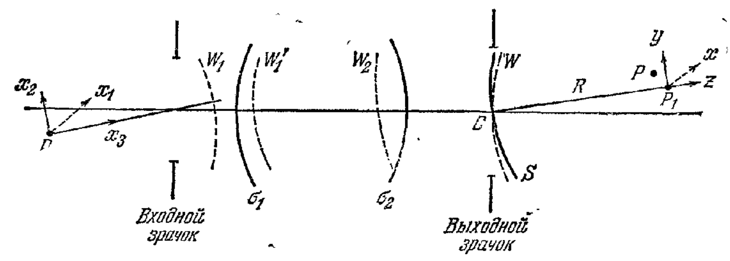


Рис. 8.7. Распространение электромагнитной волны через оптическую систему.

величина и ориентация. Запишем компоненты $\mathbf{Q}(t)$ по трем направлениям в виде интегралов Фурье, т. е.

$$Q_j(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} q_j(\omega) \exp(-i\omega t)\, d\omega \quad (j = 1, 2, 3). \tag{1}$$

Так как $Q_j(t)$ действительно, комплексные величины $q_j(\omega)$ удовлетворяют соотношениям

$$q_j(-\omega) = q_j^*(\omega), \tag{2}$$

где звездочка означает комплексное сопряжение. Следовательно, (1) можно записать в виде

$$Q_j(t) = \mathrm{Re}\left\{ \sqrt{\frac{2}{\pi}} \int_0^{\infty} q_j(\omega) \exp(-i\omega t)\, d\omega \right\} \quad (j = 1, 2, 3). \tag{3}$$

Каждая компонента Фурье в (3) представляет собой монохроматический осциллятор Герца с осью вдоль направления x_j.

Пусть $|q_j(\omega)|$ и $\delta_j(\omega)$ — амплитуда и фаза $q_j(\omega)$, т. е.

$$q_j(\omega) = |q_j(\omega)| \exp[i\delta_j(\omega)]. \tag{4}$$

Так как предполагается, что наш источник излучает квазимонохроматический свет, модуль $|q_j(\omega)|$ для каждого j заметно отличается от нуля только в узком интервале $(\omega_0 - \frac{1}{2}\Delta\omega,\ \omega_0 + \frac{1}{2}\Delta\omega)$. Предположение о том, что источник испускает естественный свет, означает, что функции $\delta_j(\omega)$ быстро и беспорядочно изменяются по частотному диапазону *).

Так как исследуемое поле можно рассматривать как суперпозицию строго монохроматических полей, то удобнее сначала исследовать вклад, вносимый в интенсивность каждым монохроматическим осциллятором Герца, находящимся в P_0. Поле такого осциллятора слабо вблизи его оси, и можно принять, что углы, под которыми из P_0 видны диаметры входного зрачка, малы; поэтому существенный вклад в поле вносят только компоненты $\mathbf{Q}(t)$, а именно $Q_1(t)$ и $Q_2(t)$. Возьмем в качестве произвольного осциллятора такой, ось которого лежит в плоскости x_1x_2.

*) Более подробное исследование этого вопроса можно найти у Планка [21].

Пусть

$$\operatorname{Re}\{q(\omega)\,\mathbf{p}_0(\omega)\exp(-i\omega t)\} \tag{5}$$

— момент такого произвольного диполя, а $\mathbf{p}_0(\omega)$ — единичный вектор в направлении оси диполя. Поле этого диполя в точке T в вакууме на расстоянии от точки P_0, большом по сравнению с длиной волны $\lambda = (2\pi c/\omega)$, определяется (см. (2.2.64)) выражениями

$$\left.\begin{aligned} \mathbf{E}_\omega &= \operatorname{Re}\left\{\frac{\omega^2}{c^2 r}\,|q(\omega)|\,\mathbf{r}_0\times(\mathbf{p}_0(\omega)\times\mathbf{r}_0)\exp\left[i\left[\delta(\omega)-\omega(t-r/c)\right]\right]\right\},\\ \mathbf{H}_\omega &= \operatorname{Re}\left\{\frac{\omega^2}{c^2 r}\,|q(\omega)|\,\mathbf{r}_0\times\mathbf{p}_0(\omega)\exp\left[i\left[\delta(\omega)-\omega(t-r/c)\right]\right]\right\}, \end{aligned}\right\} \tag{6}$$

где $\mathbf{r}_0$ обозначает единичный радиальный вектор.

Пусть W_1 — произвольный геометрический волновой фронт в пространстве предметов на расстоянии от P_0, большом по сравнению с длиной волны. Так как мы предположили, что углы, которые лучи составляют с осью системы, малы, то из (6) непосредственно следует, что в любой данный момент векторы $\mathbf{E}_\omega$ и $\mathbf{H}_\omega$ незначительно изменяются по величине и по направлению по всему фронту W_1.

Первая поверхность *) σ_1 оказывает на падающее поле двойное действие. Во-первых, амплитуды векторов поля уменьшаются вследствие потерь при отражении и, во-вторых, изменяются направления колебаний. Формулы Френеля показывают, что оба эффекта зависят главным образом от величины угла падения. Если угол мал (около 10°), потери на отражение также малы (около 5%), а поворот плоскостей колебаний не превышает нескольких градусов (см. § 1.5). Кроме того, эти эффекты практически одинаковы по всей поверхности σ_1. Так как независимые от времени части $\mathbf{E}_\omega$ и $\mathbf{H}_\omega$ незначительно изменяются по волновому фронту W_1, они столь же мало изменяются и по преломленному волновому фронту W_1', распространяющемуся после σ_1 (см. рис. 8.7). Те же рассуждения применимы к обоим полям и на любом другом волновом фронте, движущемся в пространстве между σ_1 и второй поверхностью σ_2. В самом деле, в п. 3.1.3 было показано, что в однородной среде направление колебаний вдоль каждого луча остается постоянным и, так как волновые фронты близки к сферическим (с центром в параксиальном изображении точки P_0 первой поверхностью), то амплитуды уменьшаются почти в отношении параксиальных радиусов кривизны волновых фронтов.

Повторяя те же рассуждения, мы в конце концов перейдем к волновому фронту W, проходящему через центр C выходного зрачка, и снова отметим, что независимые от времени части $\mathbf{E}_\omega$ и $\mathbf{H}_\omega$ не изменяются заметным образом по этому волновому фронту. Такой результат позволяет сразу написать приближенное математическое выражение для векторов поля в области изображения.

Пусть начало прямоугольной системы декартовых координат (x, y, z) находится в параксиальном изображении P_1 точки P_0, а ось z направлена вдоль CP_1. Поле во всех точках отверстия, за исключением точек, находящихся в непосредственной близости от его края, можно приближенно представить в виде выражений (см. гл. 3)

$$\left.\begin{aligned} &\mathbf{E}_\omega(x, y, z, t) =\\ &\quad= \operatorname{Re}\left\{\frac{\omega^3}{c^3}\,\mathbf{e}_\omega(x, y, z)\exp\left[i\left\{\delta(\omega)-\omega\left[t-\frac{1}{c}\mathcal{S}_\omega(x, y, z)\right]\right\}\right]\right\},\\ &\mathbf{H}_\omega(x, y, z, t) =\\ &\quad= \operatorname{Re}\left\{\frac{\omega^3}{c^3}\,\mathbf{h}_\omega(x, y, z)\exp\left[i\left\{\delta(\omega)-\omega\left[t-\frac{1}{c}\mathcal{S}_\omega(x, y, z)\right]\right\}\right]\right\}, \end{aligned}\right\} \tag{7}$$

*) Здесь предполагается, что σ_1 — преломляющая поверхность. Если σ_1 — зеркало, то необходимо только небольшое изменение наших рассуждений, что следует из рассмотрения формул Френеля.

которые мы вправе рассматривать как обобщение (6). Здесь $\mathcal{S}_\omega$ (x, y, z) — длина оптического пути от точки предмета до точки (x, y, z), а $\mathbf{e}_\omega$ (x, y, z) и $\mathbf{h}_\omega(x, y, z)$ — взаимно ортогональные вещественные векторы *). В однородной немагнитной среде с показателем преломления n эти векторы удовлетворяют соотношению (см. уравнения (3.1.19) и (3.1.20))

$$|\mathbf{h}_\omega| = n\,|\mathbf{e}_\omega|. \tag{8}$$

Рассмотрим опорную сферу S с центром в P_1, проходящую через центр C выходного зрачка, и обозначим ее радиус CP_1 через R. Практически расстояние между S и W нигде не превышает нескольких десятков длин волн. Следовательно, на S, так же как и на W, векторные амплитуды $\mathbf{e}_\omega$ и $\mathbf{h}_\omega$ практически постоянны по величине и по направлению.

Пусть $P(X, Y, Z)$ — точка в области изображений, где определяется интенсивность. Если углы, под которыми из P видны диаметры выходного зрачка, малы, то, используя формулу Кирхгофа с теми же допущениями, как и в предыдущем параграфе, и интегрируя выражение (7) по части S' поверхности S, приблизительно покрывающей выходной зрачок, найдем, пренебрегая изменением коэффициента наклона на S',

$$\left.\begin{aligned}\mathbf{E}_\omega(X, Y, Z, t) &= \operatorname{Re}\frac{\omega^3}{2\pi i c^3}\exp\left(i\left[\delta(\omega)-\omega t\right]\right)\iint\limits_{S'}\frac{1}{s}\,\mathbf{e}_\omega(x', y', z')\times\\ &\qquad\times\exp\left[i\frac{\omega}{c}\left[\mathcal{S}_\omega(x', y', z')+s\right]\right]dS,\\ \mathbf{H}_\omega(X, Y, Z, t) &= \operatorname{Re}\frac{\omega^3}{2\pi i c^3}\exp\left(i\left[\delta(\omega)-\omega t\right]\right)\iint\limits_{S'}\frac{1}{s}\,\mathbf{h}_\omega(x', y', z')\times\\ &\qquad\times\exp\left[i\frac{\omega}{c}\left[\mathcal{S}_\omega(x', y', z')+s\right]\right]dS.\end{aligned}\right\} \tag{9}$$

Здесь s — расстояние от произвольной точки (x', y', z') на опорной сфере до точки P.

Так как векторы $\mathbf{e}_\omega$ (x', y', z') и $\mathbf{h}_\omega$ (x', y', z') не меняются заметно по поверхности интегрирования, можно заменить их значениями $\mathbf{e}_\omega$ $(0, 0, -R)$ и $\mathbf{h}_\omega$ $(0, 0, -R)$, которые они принимают в центре C выходного зрачка. Тогда, поскольку эти векторы взаимно ортогональны и удовлетворяют (8), можно принять, если вдобавок считать $n = 1$, что

$$\left.\begin{aligned}\mathbf{e}_\omega(0, 0, -R) &= a(\omega)\,\boldsymbol{\alpha}(\omega),\\ \mathbf{h}_\omega(0, 0, -R) &= a(\omega)\,\boldsymbol{\beta}(\omega),\end{aligned}\right\} \tag{10}$$

где $\boldsymbol{\alpha}(\omega)$ и $\boldsymbol{\beta}(\omega)$ — ортогональные единичные векторы, лежащие в плоскости перпендикулярной к направлению z. Соотношение (9) принимает вид

$$\left.\begin{aligned}\mathbf{E}_\omega(X, Y, Z, t) &= \operatorname{Re}\left\{\frac{\omega^2}{c^2}U_\omega(X, Y, Z)\,a(\omega)\,\boldsymbol{\alpha}(\omega)\exp\left[i\left[\delta(\omega)-\omega t\right]\right]\right\},\\ \mathbf{H}_\omega(X, Y, Z, t) &= \operatorname{Re}\left\{\frac{\omega^2}{c^2}U_\omega(X, Y, Z)\,a(\omega)\,\boldsymbol{\beta}(\omega)\exp\left[i\left[\delta(\omega)-\omega t\right]\right]\right\},\end{aligned}\right\} \tag{11}$$

где U_ω — скалярная волновая функция вида

$$U_\omega(X, Y, Z) = \frac{\omega}{2\pi i c}\iint\limits_{S'}\frac{1}{s}\exp\left[i\frac{\omega}{c}\left[\mathcal{S}_\omega(x', y', z')+s\right]\right]dS. \tag{12}$$

Вычисляя вектор Пойнтинга $\mathbf{S}_\omega = c[\mathbf{E}_\omega\times\mathbf{H}_\omega]/4\pi$ и усредняя по времени, получим непосредственно из (11), что интенсивность в точке $P(X, Y, Z)$, обусловленная одиночным диполем (определяемым формулой (5)), помещенным

*) Вещественный характер $\mathbf{e}_\omega$ и $\mathbf{h}_\omega$ следует из того обстоятельства, что в (6) соответствующие векторы вещественны (линейная поляризация) и что поляризация остается линейной при каждом отражении. Кроме того, между любыми двумя последовательными поверхностями состояние поляризации постоянно вдоль каждого луча, как это показано в п. 3.1.3.

в P_0, пропорциональна квадрату модуля скалярной волновой функции $U_\omega(X,Y,Z)$. Однако, чтобы использование одной скалярной волновой функции для расчета интенсивности было оправдано, усреднение по времени следует выполнять не для одной монохроматической компоненты, а для всего поля.

8.4.2. Полное поле изображения. Мы видели, что вклад каждой частотной компоненты в полное поле можно рассматривать как результат действия двух диполей, помещенных в P_0, с осями, направленными вдоль x_1 и x_2. Таким образом, из (1) и (11) следует, что если мы с помощью соотношений типа (2) определим также вклад от отрицательных частот, то полное поле в области изображения приближенно можно представить в виде

$$\left.\begin{aligned}\mathbf{E}(X,Y,Z,t)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\omega^2}{c^2}U_\omega(X,Y,Z)\,[a_1(\omega)\,\boldsymbol{\alpha}_1(\omega)\exp[i\delta_1(\omega)]+\\&\qquad+a_2(\omega)\,\boldsymbol{\alpha}_2(\omega)\exp[i\delta_2(\omega)]]\exp(-i\omega t)\,d\omega,\\ \mathbf{H}(X,Y,Z,t)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\omega^2}{c^2}U_\omega(X,Y,Z)\,[a_1(\omega)\,\boldsymbol{\beta}_1(\omega)\exp[i\delta_1(\omega)]+\\&\qquad+a_2(\omega)\,\boldsymbol{\beta}_2(\omega)\exp[i\delta_2(\omega)]]\exp(-i\omega t)\,d\omega.\end{aligned}\right\}\qquad(13)$$

Здесь индексы 1 и 2 относятся к вкладам осцилляторов, оси которых направлены вдоль x_1 и x_2.

Для того чтобы определить интенсивность в области изображений, напишем отдельно выражения для каждой декартовой компоненты векторов **E** и **H**. Пусть $\theta_1(\omega)$ и $\theta_2(\omega)$ — углы, которые единичные векторы $\boldsymbol{\alpha}_1(\omega)$ и $\boldsymbol{\alpha}_2(\omega)$ составляют с осью x в пространстве изображений. Так как $\boldsymbol{\alpha}_1(\omega)$, $\boldsymbol{\beta}_1(\omega)$ и $\boldsymbol{\alpha}_2(\omega)$, $\boldsymbol{\beta}_2(\omega)$ — вещественные взаимно ортогональные векторы, лежащие в плоскости, перпендикулярной к оси z, то в соответствии с (13) получим, что компоненты векторов **E** и **H** приближенно равны *)

$$\left.\begin{aligned}E_x(X,Y,Z,t)&=H_y(X,Y,Z,t)=\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}U_\omega(X,Y,Z)\,f(\omega)\exp(-i\omega t)\,d\omega,\\ E_y(X,Y,Z,t)&=-H_x(X,Y,Z,t)=\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}U_\omega(X,Y,Z)\,g(\omega)\exp(-i\omega t)\,d\omega,\\ E_z(X,Y,Z,t)&=H_z(X,Y,Z,t)=0,\end{aligned}\right\}\qquad(14)$$

где

$$\left.\begin{aligned}f(\omega)&=\frac{\omega^2}{c^2}\,[a_1(\omega)\cos\theta_1(\omega)\exp[i\delta_1(\omega)]+a_2(\omega)\cos\theta_2(\omega)\exp[i\delta_2(\omega)]],\\ g(\omega)&=\frac{\omega^2}{c^2}\,[a_1(\omega)\sin\theta_1(\omega)\exp[i\delta_1(\omega)]+a_2(\omega)\sin\theta_2(\omega)\exp[i\delta_2(\omega)]].\end{aligned}\right\}\qquad(15)$$

Из (14) следует, что величину вектора Пойнтинга $\mathbf{S}=c[\mathbf{E}\times\mathbf{H}]/4\pi$ приближенно можно представить в виде

$$|\mathbf{S}|=\frac{c}{4\pi}\,[E_x^2+E_y^2]=\frac{c}{4\pi}\,[H_x^2+H_y^2].\qquad(16)$$

Теперь нужно определить среднее по времени этой величины **).

*) Было бы неправильно делать из (14) вывод, что в области изображения направление потока энергии обязательно параллельно z. Относительные ошибки в (14) могут существенно повлиять на определение направления в областях, где интенсивность мала, например в областях вблизи темных колец в картине Эйри.

**) Использованный здесь метод вычисления среднего предложен Борном и Йорданом [22].

Учитывая требование сходимости, мы предполагаем, что поле излучения существует только в промежутке времени между $t=-T$ и $t=T$, где $T \gg 2\pi/\omega_0$. Теперь нетрудно перейти к пределу $T \to \infty$. В силу обратной теоремы Фурье получим из (14)

$$U_\omega(X, Y, Z)\, f(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} E_x(X, Y, Z, t) \exp(i\omega t)\, dt. \tag{17}$$

Аналогичные выражения можно написать и для E_y, H_x и H_y. Тогда в соответствии с (14) имеем

$$\langle E_x^2 \rangle = \frac{1}{2T} \int_{-T}^{T} E_x^2\, dt = \frac{1}{2T} \int_{-T}^{T} E_x\, dt \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U_\omega f(\omega) \exp(-i\omega t)\, d\omega, \tag{18}$$

или, меняя порядок интегрирования и учитывая (17),

$$\langle E_x^2 \rangle = \frac{1}{2T} \int_{-\infty}^{+\infty} U_\omega f(\omega) d\omega \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} E_x \exp(-i\omega t)\, dt = \frac{1}{2T} \int_{-\infty}^{+\infty} U_\omega f_\omega U_\omega^* f_\omega^*\, d\omega =$$

$$= \frac{1}{T} \int_0^\infty |U_\omega|^2\, |f(\omega)|^2\, d\omega, \tag{19}$$

так как $U_{-\omega} = U_\omega^*$, $f(-\omega) = f^*(\omega)$. Аналогично

$$\langle E_y^2 \rangle = \frac{1}{T} \int_0^\infty |U_\omega|^2\, |g(\omega)|^2\, d\omega. \tag{20}$$

Следовательно, интенсивность $I(x, y, z)$, определенная как средняя по времени величина вектора Пойнтинга, равна, согласно (16), (19) и (20),

$$I(x, y, z) = \frac{c}{4\pi T} \int_0^\infty |U_\omega(X, Y, Z)|^2\, [\,|f(\omega)|^2 + |g(\omega)|^2]\, d\omega. \tag{21}$$

Если величина $|\Delta\omega|$ достаточно мала, то $|U_\omega|$ практически не зависит от ω в диапазоне эффективных частот, и $|U_\omega|$ можно заменить на $|U_{\omega_0}|$ и вынести за знак интеграла. Оставшийся член

$$\frac{c}{4\pi T} \int_0^\infty \{|f(\omega)|^2 + |g(\omega)|^2\}\, d\omega, \tag{22}$$

не зависящий от X, Y и Z, также не должен зависеть от T (неявно содержащегося в f и g на основании (17)), если наблюдается стационарное явление. Поэтому величина (22) постоянна (допустим равна C), и окончательно можно написать выражение для интенсивности в виде

$$I(X, Y, Z) = C\, |U_{\omega_0}(X, Y, Z)|^2. \tag{23}$$

Постоянная C сложным образом зависит от характеристик источника и оптического прибора. Однако в обычных условиях интерес представляет только относительное распределение интенсивности, а не ее абсолютная величина. В этом случае интенсивность измеряется просто величиной $|U_{\omega_0}|^2$. Таким образом, комплексная скалярная функция (12) позволяет вычислять распределение интенсивности в изображении, полученном с источником естественного света с помощью оптической системы при умеренной числовой апертуре.

§ 8.5. Дифракция Фраунгофера на отверстиях разной формы

Исследуем теперь картину дифракции Фраунгофера от отверстий разной формы.

8.5.1. Прямоугольное отверстие и щель. Рассмотрим сначала прямоугольное отверстие со сторонами $2a$ и $2b$. Если начало координат O находится в центре прямоугольника, а оси $O\xi$ и $O\eta$ параллельны его сторонам (рис. 8.8), то дифракционный интеграл Фраунгофера (8.3.36) принимает вид

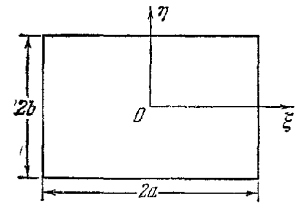


Рис. 8.8. Прямоугольное отверстие.

$$U(P) = C\int_{-a}^{a}\int_{-b}^{b}\exp\left[-ik(p\xi+q\eta)\right]d\xi\,d\eta = C\int_{-a}^{a}\exp(-ikp\xi)\,d\xi\int_{-b}^{b}\exp(-ikq\eta)\,d\eta.$$

Тогда

$$\int_{-a}^{a}\exp(-ikp\xi)\,d\xi = -\frac{1}{ikp}\left[\exp(-ikpa)-\exp(ikpa)\right] = 2\,\frac{\sin kpa}{kp}.$$

Такой же вид имеет и второй интеграл. Следовательно, интенсивность записывается в виде

$$I(P) = |U(P)|^2 = \left(\frac{\sin kpa}{kpa}\right)^2\left(\frac{\sin kqb}{kqb}\right)^2 I_0, \tag{1}$$

где, согласно (8.3.44), $I_0 = C^2D^2 = ED/\lambda^2$ — интенсивность в центре картины, E — полная энергия, падающая на отверстие, а $D = 4ab$ — площадь прямоугольника.

Таблица 8.1

Пять первых максимумов функции

$$y = \left(\frac{\sin x}{x}\right)^2$$

x	$y=\left(\frac{\sin x}{x}\right)^2$
0	1
$1,430\pi = 4,493$	0,04718
$2,459\pi = 7,725$	0,01694
$3,470\pi = 10,90$	0,00834
$4,479\pi = 14,07$	0,00503

График функции $y = (\sin x/x)^2$ приведен на рис. 8.9. Она имеет главный максимум $y = 1$ при $x = 0$ и минимумы, равные нулю, при $x = \pm\pi$, $\pm 2\pi$, $\pm 3\pi$, ... Эти минимумы разделяют вторичные максимумы, положение которых определяется корнями уравнения $\operatorname{tg} x - x = 0$ (см. табл. 8.1). Значения корней асимптотически приближаются к величинам $x = (2m+1)\pi/2$, где m — нечетное целое число.

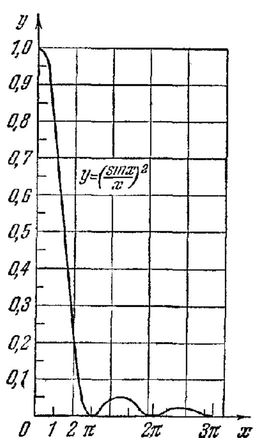


Рис. 8.9. Дифракция Фраунгофера на прямоугольном отверстии. График функции $y=\left(\frac{\sin x}{x}\right)^2$.

Очевидно, что интенсивность $I(P)$ равна нулю вдоль двух рядов линий, параллельных сторонам прямоугольника. Положение этих линий определяется из соотношений

$$kpa = \pm u\pi, \qquad kpb = \pm v\pi \qquad (u,\ v = 1,\ 2,\ 3,\ \ldots), \tag{2}$$

8.5.2. Круглое отверстие. Аналогичным образом можно исследовать дифракцию Фраунгофера на круглом отверстии. Для этого целесообразно использовать полярные координаты вместо прямоугольных. Пусть (ρ, θ) — полярные координаты произвольной точки отверстия, т. е.

$$\rho\cos\theta = \xi, \qquad \rho\sin\theta = \eta, \tag{6}$$

и пусть (w, ψ) — координаты точки P в дифракционной картине, относящейся к геометрическому изображению источника, т. е.

$$w\cos\psi = p, \qquad w\sin\psi = q. \tag{7}$$

Из определения p и q следует, что $w = \sqrt{p^2+q^2}$ равно синусу угла между направлением (p, q) и центральным направлением $p = q = 0$. В таком случае, если a — радиус круглого отверстия, то дифракционный интеграл (8.3.36) принимает вид

$$U(P) = C\int_0^a\int_0^{2\pi} \exp\left[-ik\rho w\cos(\theta-\psi)\right]\rho\,d\rho\,d\theta. \tag{8}$$

Воспользуемся теперь хорошо известным интегральным представлением бесселевых функций *) $J_n(z)$

$$\frac{i^{-n}}{2\pi}\int_0^{2\pi}\exp\left[ix\cos\alpha\right]\exp(in\alpha)\,d\alpha = J_n(x). \tag{9}$$

Тогда уравнение (8) сведется к

$$U(P) = 2\pi C\int_0^a J_0(k\rho w)\,\rho\,d\rho. \tag{10}$$

Существует также хорошо известное рекуррентное соотношение (см., например, [24] или [26])

$$\frac{d}{dx}\left\{x^{n+1}J_{n+1}(x)\right\} = x^{n+1}J_n(x), \tag{11}$$

дающее после интегрирования для $n = 0$

$$\int_0^x x'J_0(x')\,dx' = xJ_1(x). \tag{12}$$

Из (10) и (12) следует, что

$$U(P) = CD\left[\frac{2J_1(kaw)}{kaw}\right], \tag{13}$$

где $D = \pi a^2$. Следовательно, интенсивность определяется выражением

$$I(P) = |U(P)|^2 = \left[\frac{2J_1(kaw)}{kaw}\right]^2 I_0, \tag{14}$$

где в соответствии с (8.3.44) $I_0 = C^2D^2 = ED/\lambda^2$. Эта широко известная формула в несколько ином виде впервые была получена Эйри [27]**).

Распределение интенсивности в окрестности геометрического изображения описывается функцией $y = (2J_1(x)/x)^2$, график которой приведен на рис. 8.11. Она имеет главный максимум $y = 1$ при $x = 0$ и с увеличением x осциллирует с постепенным уменьшением амплитуды подобно функции $(\sin x/x)^2$,

---

*) См., например, [24] или [25].

**) Почти одновременно с Эйри, Шверд нашел приближенное решение, заменив кружок правильным многоугольником со 180 сторонами.

Векторное рассмотрение дифракции сходящейся сферической волны на круглом отверстии с учетом поляризационных особенностей поля было опубликовано в [28].

рассмотренной в п. 8.5.1. Интенсивность равна нулю (минимум) при значениях x, определяемых $J_1(x)=0$. Минимумы уже не строго эквидистантны (см. табл. 8.2).

Таблица 8.2

Несколько первых максимумов и минимумов функции

$$y=\left(\frac{2J_1(x)}{x}\right)^2$$

x	$\left(\frac{2J_1(x)}{x}\right)^2$	
0	1	Максимум
$1{,}220\pi=3{,}832$	0	Минимум
$1{,}635\pi=5{,}136$	0,0175	Максимум
$2{,}233\pi=7{,}016$	0	Минимум
$2{,}679\pi=8{,}417$	0,0042	Максимум
$3{,}238\pi=10{,}174$	0	Минимум
$3{,}699\pi=11{,}620$	0,0016	Максимум

Рис. 8.11. Дифракция Фраунгофера на круглом отверстии. График функции $y=\left(\frac{2J_1(x)}{x}\right)^2$.

Положения вторичных максимумов определяются значениями x, удовлетворяющими уравнению

$$\frac{d}{dx}\left[\frac{J_1(x)}{x}\right]=0,$$

или, пользуясь формулой *) (аналогичной (11))

$$\frac{d}{dx}\left[x^n J_n(x)\right]=-x^{-n}J_{n+1}(x), \qquad (15)$$

— корнями уравнения $J_2(x)=0$. При увеличении x расстояния между последовательными минимумами или последовательными максимумами приближаются к π, как и в предыдущем случае.

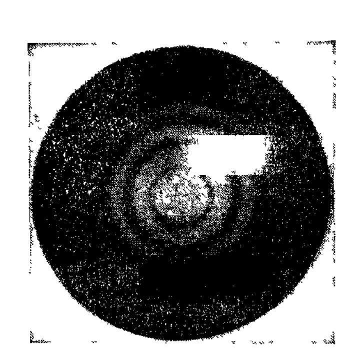


Рис. 8.12. Картина дифракции Фраунгофера на круглом отверстии (картина Эйри) диаметром 6 *мм* (по Липсону, Тейлору и Томпсону).

Увеличение 50×, желтая линия ртути $\lambda=5790$ Å. Для выделения слабых вторичных максимумов центральная часть переэкспонирована.

Найденные результаты показывают, что наблюдаемая картина имеет вид светлого диска с центром в геометрическом изображении источника $p=q=0$, окруженного светлыми и темными кольцами (рис. 8.11 и 8.12). Интенсивность светлых колец быстро уменьшается с увеличением радиуса и обычно только одно или два первых кольца достаточно ярки, чтобы их можно было наблюдать невооруженным глазом. Из табл. 8.2 следует, что, поскольку $x=2\pi a w/\lambda$, радиусы темных полос равны

$$w=\sqrt{p^2+q^2}=0{,}610\frac{\lambda}{a},\ 1{,}116\frac{\lambda}{a},\ 1{,}619\frac{\lambda}{a},\ \ldots \qquad (16)$$

Расстояние между двумя соседними кольцами асимптотически приближается к величине $\lambda/2a$. Здесь мы снова видим, что эффективные размеры дифракционной картины обратно пропорциональны линейным размерам отверстия.

*) См., например, [24] и [26].

Представляет также интерес определить, какая часть полной падающей энергии приходится на центральное пятно дифракционной картины Обозначая через $L(w_0)$ ту часть полной энергии, которая приходится на кружок радиуса w_0 в плоскости изображения с центром в геометрическом изображении, получим

$$L(w_0)=\frac{1}{E}\int_0^{2\pi}\int_0^{w_0} I(w)\,w\,dw\,d\psi=\frac{D}{\lambda^2}\int_0^{2\pi}\int_0^{w_0}\left[\frac{2J_1(kaw)}{kaw}\right]^2 w\,dw\,d\psi=2\int_0^{kaw_0}\frac{J_1^2(x)}{x}\,dx. \quad (17)$$

Учитывая уравнение (15) для $n=0$ и умножая на $J_1(x)$, получим

$$\frac{J_1^2(x)}{x}=J_0(x)\,J_1(x)-\frac{dJ_1(x)}{dx}J_1(x)=-\frac{1}{2}\frac{d}{dx}\left[J_0^2(x)+J_1^2(x)\right].$$

Поскольку $J_0(0)=1$, $J_1(0)=0$, выражение (17) сводится к

$$L(w_0)=1-J_0^2(kaw_0)-J_1^2(kaw_0). \quad (18)$$

Эта формула была получена Рэлеем [29, 30] График функции $L(w_0)$ показан на рис. 8 13. Для темных колец $J_1(kaw_0)=0$, и следовательно, часть полной энергии вне любого темного кольца просто равна $J_0^2(kaw_0)$. Для первого, второго и третьего темных колец $J_0^2(kaw_0)$ равно соответственно 0,162, 0,090 и 0,062. Таким образом, более 90% света концентрируется внутри круга, ограниченного вторым темным кольцом.

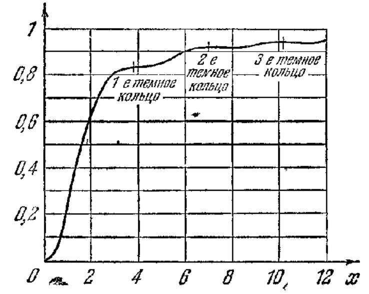


Рис 8 13 График функции $1-J_0^2(x)-J_1^2(x)$, определяющей часть полной энергии, приходящейся на кружок заданного радиуса в картине дифракции Фраунгофера на круглом отверстии

8.5.3. Отверстия другой фомы. Дифракцию Фраунгофера на отверстиях другой формы можно изучать аналогичным образом. Вычисления значительно упрощаются, если можно выбрать такие криволинейные координаты, чтобы одна из координатных осей совпадала с краем отверстия Мы не собираемся подробно рассматривать здесь другие случаи дифракции *), но выведем полезную теорему, касающуюся видоизменения дифракционной картины при симметричном расширении (или сжатии) отверстия в каком-нибудь одном направлении и, кроме того, рассмотрим дифракцию Фраунгофера на экране с большим числом отверстий одинакового размера и формы

Пусть $\mathcal{A}_1$ и $\mathcal{A}_2$ — два отверстия, причем в некотором направлении ($O\xi$) размеры $\mathcal{A}_2$ в μ раз больше, чем $\mathcal{A}_1$. Для дифракции Фраунгофера на $\mathcal{A}_1$ имеем

$$U_1(p, q)=C\iint_{\mathcal{A}_1}\exp\left[-ik(p\xi+q\eta)\right]d\xi\,d\eta. \quad (19)$$

Аналогично для дифракции Фраунгофера на $\mathcal{A}_2$ находим

$$U_2(p, q)=C\iint_{\mathcal{A}_2}\exp\left[-ik(p\xi+q\eta)\right]d\xi\,d\eta. \quad (20)$$

Заменяя здесь переменные интегрирования (ξ, η) на (ξ', η'), где

$$\xi'=\frac{1}{\mu}\xi, \qquad \eta'=\eta, \quad (21)$$

*) Дифракция Фраунгофера на кольцеобразном отверстии кратко рассматривается в связи с разрешающей силой в п 8 6 2

Фотографии картин дифракции Фраунгофера на отверстиях разной формы можно найти в [31]. Фотографии дифракционных картин Френеля были опубликованы в [32].

получим

$$U_2(p,\ q) = \mu C \iint\limits_{\mathcal{A}_1} \exp\left[-ik(\mu p\xi' + q\eta')\right] d\xi'\, d\eta' = \mu U_1(\mu p,\ q). \qquad (22)$$

Отсюда следует, что *если отверстие симметрично расширяется в каком-либо направлении в μ раз, то дифракционная картина Фраунгофера сжимается в том же направлении в μ раз, а интенсивность в некоторой точке новой картины становится в μ^2 раз больше интенсивности в соответствующей точке первоначальной картины.* Используя этот результат, можно, например, сразу найти картину дифракции Фраунгофера от эллиптического отверстия или отверстия в виде параллелограмма соответственно из картины дифракции от круга или прямоугольника. На рис. 8.14 показан случай дифракции на эллиптическом отверстии.

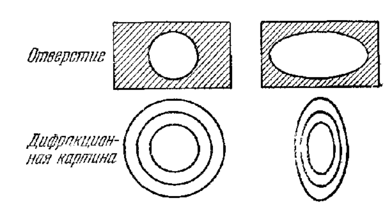


Рис. 8.14. Сравнение дифракции Фраунгофера на круглом и эллиптическом отверстиях.

Рассмотрим теперь важный случай экрана, состоящего из большого числа одинаковых и одинаково ориентированных отверстий (согласно принципу Бабине найденный результат будет также применим к дополнительному расположению препятствий). Пусть O_1, O_2, ..., O_n — ряд точек, одинаково расположенных по одной в каждом отверстии, и пусть их координаты относительно фиксированных осей, находящихся в плоскости отверстия, равны $(\xi_1,\ \eta_1)$, $(\xi_2,\ \eta_2)$, ..., $(\xi_N,\ \eta_N)$. Тогда распределение света в дифракционной картине Фраунгофера можно записать в виде

$$U(p,\ q) = C \sum_n \iint\limits_{\mathcal{A}} \exp\left[-ik\left[(\xi_n + \xi')\, p + (\eta_n + \eta')\, q\right]\right] d\xi'\, d\eta' =$$
$$= C \sum_n \exp\left[-ik\left[p\xi_n + q\eta_n\right]\right] \iint\limits_{\mathcal{A}} \exp\left[-ik(p\xi' + q\eta')\right] d\xi'\, d\eta', \qquad (23)$$

где интегрирование ведется по любому отверстию $\mathcal{A}$ из их совокупности. Интеграл выражает действие одиночного отверстия, тогда как сумма — суперпозицию когерентных дифракционных картин. Если $I^{(0)}(p,\ q)$ — распределение интенсивности, обусловленное дифракцией на одиночном отверстии, то, согласно (23), полная интенсивность равна

$$I(p,\ q) = I^{(0)}(p,\ q) \left| \sum_n \exp\left[-ik(p\xi_n + q\eta_n)\right] \right|^2 =$$
$$= I^{(0)}(p,\ q) \sum_n \sum_m \exp\left(-ik\left[p(\xi_n - \xi_m) + q(\eta_n - \eta_m)\right]\right). \qquad (24)$$

Простейший случай с двумя отверстиями рассмотрен в § 7.2 в связи с теорией интерференции. Но там мы пренебрегали зависимостью $I^{(0)}$ от p и q (т. е. влиянием дифракции на каждом отверстии) и рассматривали только эффект суперпозиции. Легко показать, что прежний результат (7.2.17) находится в согласии с соотношением (24). При $N = 2$ оно сводится к

$$I = I^0 \left\{2 + \exp\left\{-ik\left[p(\xi_1 - \xi_2) + q(\eta_1 - \eta_2)\right]\right\} + \right.$$
$$\left. + \exp\left\{-ik\left[p(\xi_2 - \xi_1) + q(\eta_2 - \eta_1)\right]\right\}\right\} = 4I^{(0)} \cos^2 \frac{1}{2}\delta,$$

где

$$\delta = k\left[p(\xi_2 - \xi_1) + q(\eta_2 - \eta_1)\right].$$

Рассмотрим теперь действие большого числа отверстий. Покажем, что результаты окажутся совершенно разными в зависимости от того, правильно ли распределены отверстия по экрану или беспорядочно.

При беспорядочном расположении отверстий члены с $m \neq n$ в двойной сумме быстро меняются в интервале от $+1$ до -1 по мере того, как p и q принимают различные значения, и сумма таких членов не достигает конечного среднего значения.

Каждый член с $m = n$ равен единице. Поэтому, если исключить локальные флуктуации *), полная интенсивность в N раз больше интенсивности света, дифрагировавшего на одном отверстии, т. е.

$$I(p, q) \approx NI^{(0)}(p, q). \tag{25}$$

Дифракционные эффекты такого рода или чаще встречающиеся дополнительные эффекты (в смысле принципа Бабине) легко наблюдаются, если, например, стеклянную пластинку, посыпанную порошком ликоподия или другим порошком с частицами одинакового размера и формы, осветить пучком света от удаленного источника. Кусок тонкой фольги, проколотый в случайно выбранных

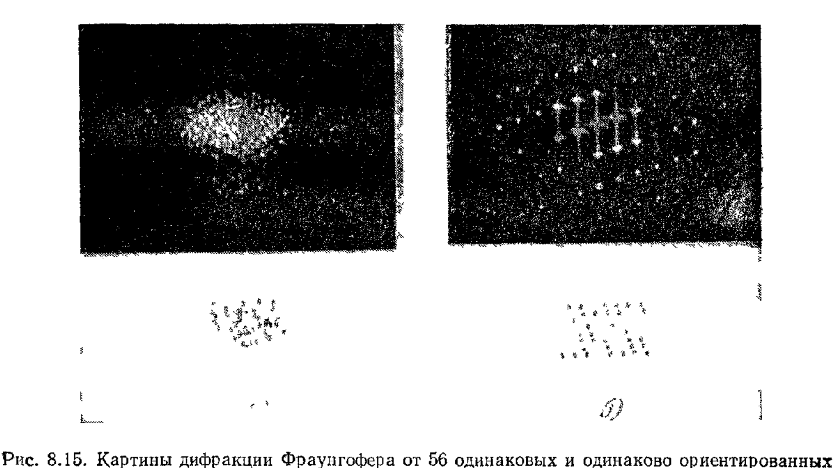


Рис. 8.15. Картины дифракции Фраунгофера от 56 одинаковых и одинаково ориентированных отверстий в плоском экране (по Липсону, Тейлору и Томпсону)

а — беспорядочная ориентировка *б* — упорядоченная ориентировка Форма и ориентировка отверстий показаны в нижней части рисунка Освещение желтой линией ртути $\lambda = 5790$ Å.

местах булавкой, также действует как дифракционный экран рассмотренного выше типа.

Совершенно иной результат получается при регулярном расположении отверстий. Тогда при некоторых значениях p и q члены с $m \neq n$ могут внести заметный вклад в интенсивность. Например, если точки O_n расположены так, что при определенных значениях p и q фазы всех членов с $m \neq n$ кратны точно 2π, то их сумма равна $N(N-1)$ и при больших N становится порядка N^2. Такое большое увеличение интенсивности в некоторых направлениях, ясно видимое на рис. 8.15, имеет, как мы увидим ниже, важное практическое значение.

---

*) Флуктуации несколько другого рода возникают, если отверстия имеют не совсем одинаковую форму, но расположены регулярно или согласно какому-то статистическому закону (см. [33]). Аналогичные эффекты наблюдаются при дифракции рентгеновских лучей в жидкостях (см. [34]).

§ 8.6. Дифракция Фраунгофера в оптических приборах

8.6.1. Дифракционные решетки. *а) Принцип действия дифракционной решетки.* Дифракционной решеткой можно назвать любое устройство, обеспечивающее периодическую модуляцию падающей волны по амплитуде или фазе, или одновременно по обоим этим параметрам. Любую данную решетку можно охарактеризовать ее *функцией пропускания*, определение которой мы сейчас дадим.

Пусть прозрачный или полупрозрачный предмет (не обязательно периодический) закрывает часть воображаемой координатной плоскости $\xi\eta$ и освещается плоской монохроматической волной, направление падения которой определяется косинусами l_0 и m_0. На рис. 8.16, иллюстрирующем это расположение, ось η перпендикулярна к плоскости рисунка. В отсутствие предмета возмущение в плоскости $\xi\eta$ описывается функцией $V_0(\xi, \eta) = A \exp\{ik(l_0\xi + m_0\eta)\}$; множитель $\exp(-i\omega t)$, как обычно, здесь опущен. В присутствии предмета возмущение изменяется и описывается некоторой другой функцией, например $V(\xi, \eta)$. Тогда *функция пропускания* предмета определяется соотношением

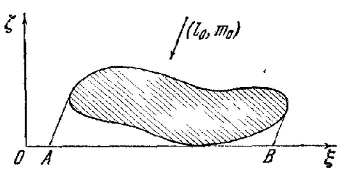


Рис. 8.16. К определению функции пропускания.

$$F(\xi, \eta) = \frac{V(\xi, \eta)}{V_0(\xi, \eta)}. \qquad (1)$$

Вообще говоря, функция пропускания F зависит не только от ξ и η, но и от направления (l_0, m_0) освещения; кроме того, в общем случае функция F комплексна, так как при прохождении сквозь предмет свет может менять и амплитуду, и фазу. В частности, когда предмет изменяет только амплитуду падающей волны, не изменяя ее фазу (т. е. если $\arg F \equiv 0$), мы говорим об *амплитудном объекте*, если же он изменяет фазу, а не амплитуду (т. е. $|F| = 1$), мы говорим о *фазовом объекте*.

Если изучается свет, отраженный предметом, а не прошедший сквозь него, то рациональнее говорить о *функции отражения*, определяемой так же, как и функция пропускания, но тогда координатная плоскость помещается с той же стороны предмета, с которой падает свет.

Отношение $|V/V_0|$ практически равно единице для всех точек, находящихся вне геометрической тени (граница ее на рис. 8.16 обозначена точками A и B), отбрасываемой предметом. Закрывая область, лежащую за пределами тени, непрозрачным экраном, мы получим устройство, действующее как дифракционное отверстие $\mathcal{A}$ с неоднородной функцией зрачка (см. (8.3.39)). Пусть линейные размеры $\mathcal{A}$ велики по сравнению с длиной волны, и пусть функция F остается в достаточной мере постоянной в областях, размеры которых сравнимы с длиной волны света; тогда дифракционная формула (8.3.23) остается справедливой при тех же условиях, что и раньше, если подынтегральное выражение дифракционного интеграла умножено на F.

Рассмотрим одномерную решетку, состоящую из N параллельных штрихов произвольного профиля, нанесенных на одну из поверхностей плоскопараллельной стеклянной пластинки. Пусть плоскость $\xi\eta$ совпадает с плоской поверхностью пластинки, η — направление штрихов, а d — период решетки в направлении ξ (рис. 8.17).

Предположим, что направление падающей на решетку волны лежит в плоскости рисунка и образует угол θ_0 с осью $O\zeta$, и пусть θ — угол между осью $O\zeta$ и линией, соединяющей очень удаленную точку наблюдения P

с решеткой. Как и раньше, положим $l_0 = \sin\theta_0$, $l = \sin\theta$, $p = l - l_0 = \sin\theta - \sin\theta_0$. В таком случае из (8.5.23) сразу же получаем выражение для комплексной амплитуды в P, если подынтегральное выражение умножается на

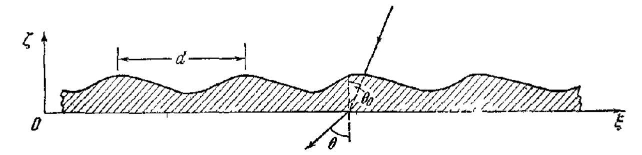


Рис. 8.17. К теории дифракционной решетки.

функцию пропускания F одного периодического элемента. Полагая $q = 0$ и

$$\xi_n = nd, \quad \eta_n = 0 \quad (n = 0, 1, \ldots, N-1), \tag{2}$$

находим

$$U(p) = U^{(0)}(p) \sum_{n=0}^{N-1} \exp(-iknd\,p) = U^{(0)}(p) \frac{1 - \exp(-iNk\,d\,p)}{1 - \exp(-ik\,d\,p)}, \tag{3}$$

где *)

$$U^{(0)}(p) = C \int_{\mathcal{A}} F(\xi) \exp(-ikp\xi)\, d\xi. \tag{4}$$

Следовательно,

$$I(p) = |U(p)|^2 = \frac{(1 - \exp(-iNk\,dp))}{(1 - \exp(-ik\,dp))} \cdot \frac{(1 - \exp(iNk\,dp))}{(1 - \exp(ik\,dp))} |U^{(0)}(p)|^2 = \frac{1 - \cos Nk\,dp}{1 - \cos k\,dp} I^{(0)}(p), \tag{5}$$

где $I^{(0)}(p) = |U^{(0)}(p)|^2$. Если ввести функцию

$$H(N, x) = \left(\frac{\sin Nx}{\sin x}\right)^2, \tag{6}$$

то формула (5) для интенсивности примет вид

$$I(p) = H\left(N, \frac{k\,dp}{2}\right) I^{(0)}(p). \tag{5a}$$

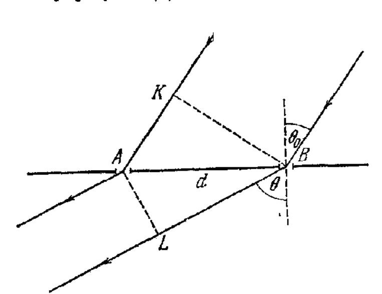


Рис. 8.18. К теории дифракционной решетки.

Прежде чем обсуждать содержание этой основной формулы, отметим, что, согласно (3), распределение интенсивности будет таким же, как и распределение, обусловленное системой вторичных когерентных источников; каждый из них характеризуется одной и той же амплитудной функцией $|U^{(0)}(p)|$, а их фазы отличаются на величину, кратную kdp. Чтобы понять значение этой разности фаз, рассмотрим две соответствующие точки A и B в соседних штрихах решетки (рис. 8.18). Так как влияние решетки заключается в периодическом воздействии на падающую волну, то разность хода между световыми лучами, приходящими в A и B, такая же, как и в отсутствие решетки, т. е. равна $AK =$

*) Так как F зависит от l_0, то величины $U^{(0)}$ и $I^{(0)}$ зависят и от l и от l_0, а не только от разности $l - l_0$. Поскольку нас интересуют только эффекты, возникающие при фиксированном направлении падения света, то мы можем считать l_0 постоянным и сохранить предыдущие обозначения.

$=d\sin\theta_0$. Здесь точка K — основание перпендикуляра, опущенного из B на луч, падающий в A. Далее путь светового луча, выходящего из B в направлении θ, превышает путь луча, выходящего из A в том же направлении, на $BL=d\sin\theta$; здесь L — основание перпендикуляра, опущенного из A на луч, дифрагировавший в B в направлении θ. Следовательно, полная разность хода между светом, приходящим в удаленную точку наблюдения от соответствующих точек двух соседних штрихов, равна

$$BL-AK=d(\sin\theta-\sin\theta_0)=dp, \tag{7}$$

а соответствующая разность фаз равна $2\pi dp/\lambda=kdp$.

Формула (5а) выражает $I(p)$ в виде произведения двух функций. Одна из них $I^{(0)}$ определяет действие одного периода решетки, другая H отражает эффект интерференции света от различных периодов решетки. Функция $H(N, x)$ имеет максимумы величиной N^2 во всех точках, в которых исчезает знаменатель $\sin^2 x$, т. е. в тех точках, где x равно нулю или целому кратному π. Следовательно, функция $H(N, kdp/2)$ имеет максимумы величиной N^2, когда

$$p\equiv\sin\theta-\sin\theta_0=\frac{m\lambda}{d} \tag{8}$$

$$(m=0,\ \pm1,\ \pm2,\ \ldots).$$

Целое число m, согласно (7), представляет собой разность хода (в длинах волн) между светом, дифрагирующим в направлении максимумов и распространяющимся от соответствующих точек двух соседних штрихов. В согласии с нашим прежним определением (см. п. 7.3.1) назовем m *порядком интерференции*. В промежутках между этими главными максимумами находятся слабые вторичные максимумы (рис. 8.19, *а*). Когда N велико, первый вторичный максимум составляет лишь несколько процентов от главного максимума. Максимумы разделяются точками с нулевой интенсивностью при $x=kdp/2=\pm n\pi/N$, т. е. расположены в направлениях, определяемых соотношениями

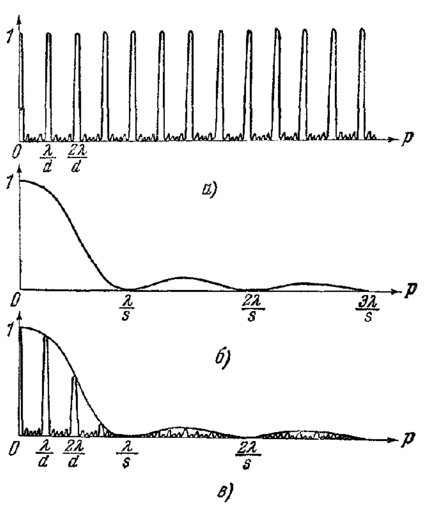


Рис. 8.19. Графики интерференционной функции и функции интенсивности.

а) — нормированная интерференционная функция $\frac{1}{N^2}H(N, k\,dp/2)=\left[\frac{\sin Nk\,d\,p/2)}{N\sin(k\,d\,p/2)}\right]^2$, *б* — нормированная функция интенсивности в случае щели $I^{(0)}(p)=\left[\frac{\sin(ksp/2)}{ksp/2}\right]^2$; *в* — нормированная функция интенсивности в случае решетки, состоящей из N одинаковых эквидистантных параллельных щелей $\frac{1}{N^2}I(p)=\left[\frac{\sin(Nkdp/2)}{N\sin(kdp/2)}\right]^2\left[\frac{\sin(ksp/2)}{ksp/2}\right]^2$. Показана только область $p\geqslant0$ все кривые симметричны относительно вертикальной оси $p=0$.

$$p\equiv\sin\theta-\sin\theta_0=\frac{n\lambda}{Nd} \qquad (n=\pm1,\ \pm2,\ \ldots); \tag{9}$$

случай, когда n/N равно целому числу, исключается.

Функция $I^{(0)}(p)$ зависит от формы штрихов. Предположим, что она имеет главный максимум в некотором направлении $p=p'$ и что с обеих сторон максимума она спадает медленно по сравнению с изменением H. В таком случае

$I(p)$ имеет общий вид интерференционной функции H, но «модулированной» функцией $I^{(0)}$. Следовательно, у $I(p)$ также существуют достаточно острые максимумы в направлении, близком к $p=m\lambda/d$. Так как эти направления (исключая случай $m=0$) зависят от длины волны, то очевидно, что решетка разлагает пучок немонохроматического света в *спектры нескольких порядков*.

Для иллюстраций изложенного выше рассмотрим решетку, состоящую из ряда длинных эквидистантных щелей (рис. 8.20) шириной s и длиной L

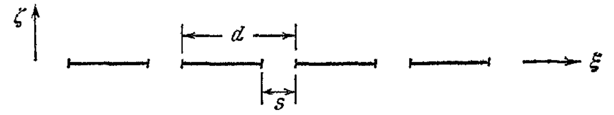


Рис. 8.20. Простая линейная решетка.

в непрозрачном экране. Если решетка освещается очень удаленным линейным источником, расположенным параллельно щелям, то интенсивность $I^{(0)}$ определяется выражением (8.5.4) (с $2a=s$, $2b=L$), и мы получим

$$I(p)=\frac{s^2L}{\lambda}\left(\frac{\sin(Nk\,dp/2)}{\sin(k\,dp/2)}\right)^2\left(\frac{\sin(ksp/2)}{ksp/2}\right)^2. \qquad (10)$$

На рис. 8.19 приведены кривые, соответствующие обоим множителям в (10) и их произведению. Последний множитель в (10), представляющий действие одиночной щели, имеет главный максимум при $p=0$ и минимумы при $ksp/2=m\pi$, т. е. при

$$p=n\lambda/s \qquad (n=\pm 1,\ \pm 2,\ \ldots) \qquad (11)$$

разделенные слабыми вторичными максимумами. Если $\lambda/s \gg \lambda/d$, т. е. если ширина каждой щели мала по сравнению с d, то в дополнение к главному максимуму при $p=0$ функция $I(p)$ имеет с обеих сторон от него ряд резких, но постепенно убывающих максимумов, расположенных в направлениях, определяемых (8).

Возвращаясь к общему случаю, мы видим, что при очень малой ширине каждого штриха — порядка длины волны (практически это часто бывает), выведенная в приближении Кирхгофа формула (4) не будет справедлива. В таких случаях для установления детального распределения интенсивности требуется более тонкое исследование. Однако мы вправе ожидать, что основные качественные характеристики, указанные нашей элементарной теорией, а именно существование резких максимумов, положение которых по существу определяется интерференционной функцией H, остаются правильными даже при очень тонких штрихах; конечно, при этом функция интенсивности для единичного периода должна мало изменяться в интервале порядка $\Delta p \sim \lambda/d$.

Выясним теперь, какова разрешающая сила дифракционной решетки. Расстояние между главным максимумом порядка m и соседним минимумом определяется, согласно (9), как

$$\Delta p=\frac{\lambda}{Nd}. \qquad (12)$$

Если длина волны изменится на величину $\Delta\lambda$, максимум порядка m сместится, согласно (8), на величину

$$\Delta' p=\frac{|m|}{d}\Delta\lambda. \qquad (13)$$

Предположим, что линии с длинами волн $\lambda\pm\frac{1}{2}\Delta\lambda$ начинают разрешаться, если максимум одной длины волны совпадает с минимумом другой (см. стр. 307); тогда для предела разрешения в порядке m получим $\Delta p=\Delta' p$, т. е.

$$\frac{\lambda}{\Delta\lambda}=|m|N. \qquad (14)$$

Следовательно, *разрешающая сила равна произведению номера порядка m на число штрихов N*. Для порядка m, согласно (8), $d(\sin\theta - \sin\theta_0) = m\lambda$, и разрешающую силу можно записать в виде

$$\frac{\lambda}{\Delta\lambda} = \frac{Nd\,|\sin\theta - \sin\theta_0|}{\lambda}. \tag{14a}$$

В соответствии с (7) это означает, что *разрешающая сила равна числу длин волн, укладывающихся в разность хода двух лучей, дифрагировавших от краев решетки (расстояние между ними составляет Nd) в направлении θ*. Следует отметить, что поскольку $|\sin\theta - \sin\theta_0|$ не может превысить 2, разрешающая сила решетки шириной w не может быть больше $2w/\lambda$.

Для иллюстрации формулы (14) определим число штрихов, которое должна иметь решетка, чтобы разрешить две линии, отстоящие друг от друга на 0,1 Å и лежащие в центре видимой области спектра. В этом случае $\lambda \approx 5500$ Å, $\Delta\lambda = 10^{-1}$ Å, и при наблюдении во втором порядке ($m = 2$), согласно (14), получим $N \geqslant 5{,}5\cdot 10^3/2\cdot 10^{-1} = 27\,500$. Таким образом, решетка должна иметь по крайней мере 27 500 штрихов.

Для сравнения рассмотрим разрешающию силу призмы в положении наименьшего отклонения, освещаемую линейным источником, параллельным ребру A (щель спектрографа). Пучок параллельных лучей, падающий на призму, дифрагирует на прямоугольнике шириной $l_1 = l_2$ (см. рис. 4.28). Согласно (8.5.2) первый минимум интенсивности находится на угловом расстоянии (предполагается, что оно мало)

$$p = \lambda/l_1 \tag{15}$$

от геометрического изображения щели. Изменение угловой дисперсии, соответствующее изменению длины волны на величину $\Delta\lambda$, согласно (4.7.36), равно

$$\Delta\varepsilon = \frac{t}{l_1}\frac{dn}{d\lambda}\Delta\lambda, \tag{16}$$

где t — наибольшая толщина стекла, сквозь которую проходит один из крайних лучей, а n — показатель преломления материала призмы. На пределе разрешения $p \approx \Delta\varepsilon$, и поэтому разрешающая сила определяется соотношением

$$\frac{\lambda}{\Delta\lambda} = t\left|\frac{dn}{d\lambda}\right|. \tag{17}$$

Последнее соотношение показывает, что *при данном сорте стекла разрешающая сила призмы зависит только от наибольшей толщины стекла* *), *которую лучи проходят в призме; в частности, разрешающая сила не зависит от угла призмы.* Например, пусть длина основания призмы равна 5 *см* и сделана она из тяжелого флинта с $dn/d\lambda \sim 1000\ см^{-1}$ на длине волны $\lambda = 5500$ Å. Если работает вся призма, то, согласно (17), в средней части видимой области спектра разрешаются линии, находящиеся на расстоянии, не меньшем $\Delta\lambda$, где $\lambda \sim 5{,}5\cdot 10^{-5}\ см/5{,}10^3 = 1{,}1$ Å. Следовательно, у такой призмы довольно больших размеров разрешающая сила в десять раз меньше, чем у решетки с 27 500 штрихами, о которой мы говорили выше.

До сих пор мы рассматривали только одномерную решетку, но наш анализ легко распространить на дву- и трехмерную периодические структуры, вызывающие дифракцию. Двумерные решетки (называемые пересекающимися) не нашли практического применения, хотя действие подобных решеток наблюдается часто, например при рассматривании яркого источника света сквозь тонкую ткань (хотя бы носовой платок). Теория же трехмерных решеток представляет огромный практический интерес, так как такие решетки образуются в кристаллах благодаря правильному расположению атомов. Постоянная этой

*) По существу разрешающая сила призмы зависит не от наибольшей толщины стекла, а от разности в длинах верхнего и нижнего оснований. (*Прим. перев.*)

решетки (расстояние между соседними атомами) порядка 1 Å (10^{-8} *см*), что совпадает по порядку величины с длиной волн рентгеновских лучей. Следовательно, при прохождении сквозь кристалл пучка рентгеновских лучей возникает дифракционная картина, и, анализируя ее, можно получить информацию относительно структуры кристалла (см., например, [35] или [36]). Дифракция рентгеновских лучей в кристаллах, предсказанная в 1912 г. Лауэ, наблюдалась впервые Фридрихом и Книппингом [37].

Другим примером решеткоподобной структуры служат ультразвуковые волны в жидкости. Они представляют собой упругие волны, создаваемые пьезоэлектрическим генератором, и отличаются от обычных звуковых волн только значительно более высокой частотой, лежащей существенно выше верхнего предела слышимости. Такие волны вызывают периодические разряжения и сжатия жидкости, действующие на проходящий свет, как решетка. Теория этого явления излагается в гл. 12. В заключение настоящего параграфа рассмотрим одномерные решетки, которые применяются в спектроскопии.

б) Типы дифракционных решеток *). Принцип действия дифракционной решетки был открыт в 1785 г. Риттенхаузом ([39], см. также [40]), но тогда его открытие не привлекло внимания, и решетка была изобретена вновь в 1819 г. Фраунгофером [41]. Первые решетки, изготовленные Франугофером, были выполнены из очень тонкой проволоки, навитой на два параллельных винта. Изготовление таких проволочных решеток сравнительно несложно, и поэтому в некоторых случаях ими пользуются даже в наше время, особенно в длинноволновой (инфракрасной) области спектра. Позднее Фраунгофер с помощью специальной машины прочерчивал штрихи на золотой пленке, нанесенной на стеклянную пластинку, или же, пользуясь алмазным резцом,— параллельные штрихи на самой поверхности стекла.

Значительных успехов в технике изготовления решеток достиг Роуланд [42], который создал несколько превосходных гравировальных машин и, кроме того, изобрел так называемую вогнутую решетку (см. стр. 378, 379). При помощи машины Роуланда можно было наносить штрихи длиной более 10 *см* на решетку шириной 15 *см*. Первая машина Роуланда позволяла прочерчивать около 560 штрихов на 1 *мм*, что обеспечивало разрешающую силу решетки, превышающую 150 000. Позднее Майкельсон изготовлял решетки значительно шире 15 *см* с разрешающей силой, примерно равной 400 000.

Большинство первых решеток гравировались на зеркальном металле или стекле, позднее появились решетки со штрихами, нанесенными на напыленный слой алюминия. Алюминий обладает хорошей отражательной способностью в ультрафиолете и, кроме того, вследствие своей мягкости не вызывает быстрого износа чертящего острия (алмаза).

У идеальной решетки штрихи должны быть строго параллельны и иметь одинаковую форму. Однако на практике оба эти требования обычно не выполняются. Совершенно нерегулярные ошибки приводят к затуманиванию картины и не так серьезны, как систематические ошибки, например в периоде решетки. Последние приводят к появлению в спектре паразитных линий, называемых *духами*. Очень часто такие духи с большим трудом можно отличить от настоящих линий.

Высокая разрешающая сила — не единственное серьезное требование, предъявляемое практической спектроскопией. В тех случаях, когда энергии недостаточно (с чем приходится сталкиваться, например, при изучении спектров слабых звезд и туманностей или при работе в инфракрасной области спектра) существенно, чтобы как можно больше света дифрагировало в определенном направлении. Кроме того, при точных измерениях длин волн применяемая ре-

*) Подробное изложение методов изготовления решеток и их усовершенствования см. в [38].

шетка должна обладать большой дисперсией. Согласно (8) угловая дисперсия (при постоянном угле падения света) равна

$$\frac{d\theta}{d\lambda}=\frac{1}{\cos\theta}\frac{m}{d};\qquad(18)$$

следовательно, для получения большой дисперсии расстояние d должно быть малым или наблюдения должны вестись в высоких порядках (большое m). При работе с решетками, имеющими вид последовательности непрозрачных и прозрачных (или отражающих) полосок, только малая часть падающего света попадает в какой-нибудь один порядок. Это недостаток устраняется у современных решеток тем, что наносимым штрихам придают определенную форму Специальная решетка, форма штрихов которой показана на рис. 8.21, позволяет направить большую часть света в один или два порядка, лежащие по одну сторону от центрального изображения. Решетки такого типа с довольно грубыми штрихами называются *эшелетами*, так как они занимают промежуточное положение между решетками более старых типов и так называемыми эшелонами, которые будут описаны ниже. Эшелеты были впервые изготовлены Вудом [43] на медных пластинках. В качестве резца он пользовался естественной гранью специально отобранного кристалла карборунда. Позднее такие решетки гравировались алмазным острием, которому шлифовкой придавалась нужная форма. Эти решетки имели от 80 до 120 штрихов на 1 *мм*; при работе с видимым светом они посылают большую его часть в спектры 15-го или 30-го порядка. Эшелеты представляют значительную ценность для инфракрасной спектроскопии.

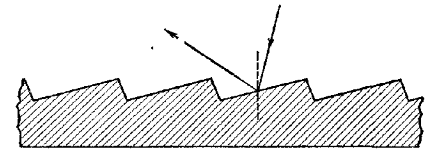


Рис. 8.21. Отражательная решетка с заданной формой штрихов.

Сравнительно недавно были разработаны методы нанесения штрихов заданного профиля на решетки со значительно меньшим расстоянием между штрихами [44]. Такие *«блестящие» решетки* *), как их называют, со штрихами, похожими по форме на штрихи эшелета, образуют наиболее интенсивные спектры в очень низких порядках (обычно в первом или во втором).

По-видимому, разрешающая сила решеток описанных типов достигает примерно 400 000 и ограничивается практическими трудностями изготовления. В некоторых случаях (например, при изучении эффекта Зеемана и дифракционной картины сверхтонкой или изотопической структуры) необходима разрешающая сила, превосходящая эту величину. Такую высокую разрешающую силу имеют предложенные Гаррисоном [45] решетки, называемые эшелями. Эти решетки с широкими и неглубокими штрихами рассчитывают для работы при углах падения света, больших 45°, причем направление падения должно быть нормальным к узкой стороне ступеньки. При работе с эшелями используются сравнительно высокие порядки ($m\sim 1000$). Эшель длиной 25 *см* с 4 штрихами на 1 *мм*, рассчитанная для работы в 1000-м порядке, имеет разрешающую силу **) в 1 000 000.

Изготовление решеток хорошего качества представляет значительные трудности, и поэтому часто используются *реплики* ***). Они изготовляются путем отливки матрицы с оригинальной гравированной решетки.

---

*) В русской литературе такие решетки называются концентрирующими. (*Прим. ред.*)

**) Теория и методы изготовления решеток с высокой разрешающей силой приведены в [46].

***) Первые реплики были получены Торпом [47], а затем Уоллесом [48]. Улучшенные способы изготовления реплик описаны в [49].

Необходимо упомянуть еще об одной «решетке» совершенно другого типа, а именно об эшелоне Майкельсона [50]. Он состоит из ряда совершенно одинаковых плоскопараллельных стеклянных пластинок, собранных в виде ступеней лестницы (отсюда и название), как показано на рис. 8.22. Каждая ступенька задерживает идущий сквозь нее пучок света, причем время задержки одного пучка относительно соседнего постоянно. Так как ширина каждой ступеньки велика по сравнению с длиной волны, влияние дифракции ограничено малыми углами, и большая часть света концентрируется в одном или двух спектрах вблизи направления $\theta = 0$, соответствующих очень высоким порядкам, поскольку разность хода между соседними пучками достигает очень большого числа длин волн.

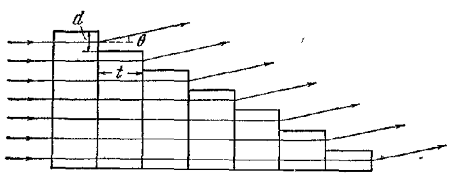


Рис. 8.22. Эшелон Майкельсона.

Разрешающая сила эшелона Майкельсона зависит не только от разности хода между лучами от краев решетки, но и от дисперсии стекла (хотя и в значительно меньшей степени). Если n — показатель преломления стекла, t — толщина каждой ступеньки и d — их ширина (см. рис. 8.22), то, считая p малым, находим, что разность хода между лучами, дифрагирующими от соседних ступеней, равна, очевидно, $pd + (n-1)t$. Отсюда для положения главных максимумов получим

$$pd + (n-1)t = m\lambda \qquad (m = 0,\ 1,\ 2,\ \ldots). \tag{19}$$

Если длина волны изменится на величину $\Delta\lambda$, то максимум порядка m сместится на

$$\Delta' p = \left| m - t\frac{dn}{d\lambda} \right| \frac{\Delta\lambda}{d}. \tag{20}$$

Расстояние Δp между главным максимумом порядка m и соседним минимумом снова определяется соотношением (12), так что условие $\Delta p = \Delta' p$ для предела разрешения дает

$$\frac{\lambda}{\Delta\lambda} = N\left| m - \frac{dn}{d\lambda}t \right|. \tag{21}$$

Здесь m можно заменить величиной $(n-1)t/\lambda$, полученной из (19), отбрасывая член pd; в самом деле, величины p, для которых интенсивность значительна, порядка λ/d, т. е. pd порядка длины волны, тогда как $(n-1)t$ порядка многих тысяч длин волн. Отсюда находим следующее выражение для разрешающей силы эшелона:

$$\frac{\lambda}{\Delta\lambda} \approx N\left| \frac{n-1}{\lambda} - \frac{dn}{d\lambda} \right| t. \tag{22}$$

Отношение $\frac{dn}{d\lambda} \Big/ \left(\frac{n-1}{\lambda}\right)$ мало. Для флинта его величина вблизи середины видимой области лежит в интервале от $-0{,}05$ до $-0{,}1$. Следовательно, при наблюдении в порядке $m = (n-1)t/\lambda$ разрешающая сила эшелона в этих условиях превосходит на 5—10% разрешающую силу линейной решетки с N штрихами. Один из эшелонов Майкельсона состоял из 20 пластинок, каждая толщиной $t = 18$ *мм* и шириной $d = 1$ *мм*. Считая $n = 1{,}5$, найдем для разности хода между соседними пучками, измеренной в длине волны зеленого света $\lambda = 5{,}10^{-5}$ *см*, величину $m \approx 0{,}5 \cdot 1{,}8/5 \cdot 10^{-5} \approx 20\,000$. Полагая $\frac{dn}{d\lambda} \Big/ \left(\frac{n-1}{\lambda}\right) = -0{,}1$, получим разрешающую силу, примерно равную $20(20\,000 + 0{,}1 \cdot 20\,000) = 440\,000$.

Более важное значение имеет *отражательный эшелон*. У него каждая ступенька покрывается высоко отражающим металлическим слоем и спектры наблюдаются в отраженном свете. Разрешающая сила отражательного эшелона превосходит разрешающую силу пропускающего эшелона таких же размеров в 3—4 раза, так как каждая ступенька вносит запаздывание между соседними пучками на величину $2t$ вместо $(n-1)t \approx t/2$. Так же, как и у эшели, у отражательного эшелона разрешающая сила может превышать миллион. Другое преимущество отражательного эшелона перед пропускающим состоит в том, что его можно использовать в ультрафиолетовой области спектра, где стекло непрозрачно. Майкельсон отчетливо представлял преимущества, которые имеет отражательный эшелон перед пропускающим, но технические трудности долго препятствовали изготовлению хороших приборов такого типа. Только спустя почти тридцать лет эти трудности были успешно преодолены Вильямсом [51]. Серьезные трудности представляет точная сборка эшелона, состоящего из большого числа пластинок одинаковой толщины. Поэтому число ступенек в эшелонах практически не превышает 40.

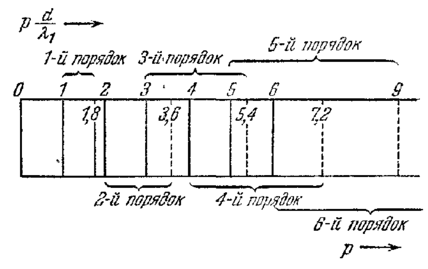


Рис. 8.23. Перекрывание спектров разных порядков, полученных с решеткой.

Наконец надо сделать несколько замечаний относительно перекрывания спектров, принадлежащих разным порядкам. Ограничимся рассмотрением видимого спектра, т. е. области длин волн от $\lambda_1 = 0{,}4$ *мк* до $\lambda_2 = 0{,}75$ *мк*. Мы видим, что спектр первого порядка не налагается на спектр второго порядка, так как спектр первого порядка занимает участок от $p = \lambda_1/d$ до $p = \lambda_2/d = 0{,}75\,\lambda_1/0{,}4d = 1{,}8\lambda_1/d$, тогда как спектр второго порядка начинается при $p = 2\lambda_1/d$. Вместе с тем спектр второго порядка занимает область от $p = 2\lambda_1/d$ до $p = 2\lambda_2/d$, а спектр третьего порядка начинается уже при $2\cdot 1{,}8\lambda_1/d$, так что оба они частично покрываются. При увеличении порядка соседние спектры перекрываются все больше и больше (рис. 8.23). Если линии с длинами волн λ и $\lambda + \delta\lambda$ налагаются в двух соседних порядках $(m+1)$ и m, то

$$(m+1)\lambda = m(\lambda + \delta\lambda), \quad \frac{\delta\lambda}{\lambda} = \frac{1}{m}. \tag{23}$$

Таким образом, *свободная область дисперсии обратно пропорциональна порядку спектра*.

Перекрывание порядков в свое время применялось для сравнения длин волн в методе совпадений (см. стр. 313). Позднее для этой цели начали пользоваться простой интерполяцией между длинами волн эталонных линий, которые определялись интерферометрически.

В заключение подведем итог главным отличительным чертам решеток различных типов. Напомним, что, согласно (14), высокая разрешающая сила достигается либо с большим числом периодов и относительно малыми порядками, либо с умеренным числом периодов и высокими порядками. Обычно с гравированными решетками работают в очень низких порядках (m от 1 до 5), тогда как эшелоны используются в высоких порядках ($m \approx 20\,000$). В промежутках находятся эшелеты (m от 15 до 30) и эшели ($m \approx 1000$). Для практических целей следует помнить, что угловая дисперсия прямо пропорциональна порядку спектра и обратно пропорциональна периоду, а свободная область дисперсии обратно пропорциональна порядку.

в) Спектрографы с дифракционной решеткой. В спектрографах с дифракционной решеткой цветное изображение щелевого источника воспроизводится в различных порядках, в которые решетка разлагает падающий свет. Простое устройство такого спектрографа показано на рис. 8.24. Коллимированный свет от щелевого источника S, находящегося в фокальной плоскости линзы L, падает на отражательную решетку G. Изображения S, образованные дифрагировавшими лучами, наблюдаются в фокальной плоскости F зрительной трубы T. Литров предложил более компактное автоколлимационное устройство

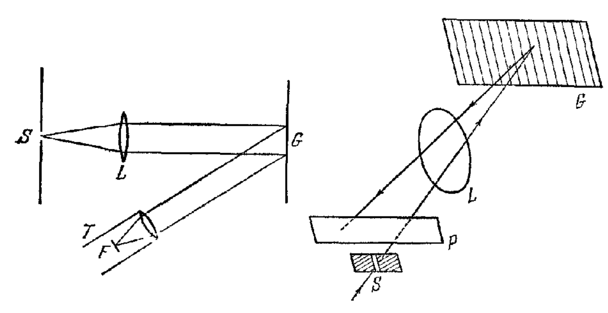


Рис. 8.24. Дифракционный спектрограф с решеткой.

Рис. 8.25 Дифракционный спектрограф по схеме Литрова.

Рис. 8.26 Фокусировка в случае вогнутой решетки (круг Роуланда).

(рис. 8.25), работающее только с одной линзой. В этом устройстве щель находится непосредственно под пластинкой P, а линза расположена вблизи решетки, которая может поворачиваться на заданный угол относительно падающего пучка.

Для устранения потерь света, возникающих при фокусировке линзой дифрагировавших лучей, Роуланд ввел в практику *вогнутую решетку*. В такой решетке штрихи гравируются на вогнутой хорошо отражающей поверхности металла, т. е. на поверхности вогнутого зеркала. Штрихи располагают так, чтобы их проекции на хорду зеркальной поверхности были эквидистантны. Простая теорема геометрии определяет возможные взаимные положения щели и плоскости наблюдения относительно решетки.

Предположим, что Q — средняя точка поверхности решетки, а C — центр ее кривизны, и опишем окружность K с центром в точке O на QC и радиусом $r = OQ = OC$ (рис. 8.26). Докажем теперь, что свет, идущий от любой точки S окружности K, отражается приблизительно в точку P, а дифрагировавшие лучи собираются в точках P', P'', ..., лежащих на нашей окружности, причем в каждую из них фокусируются дифрагировавшие лучи определенного порядка. Чтобы доказать это, построим отраженный луч QP, соответствующий падающему лучу SQ. Если $\alpha = SQC$ — угол падения, то угол отражения CQP также равен α; и, кроме того, дуга SC равняется дуге CP. Рассмотрим теперь другой луч, выходящий из S и падающий на решетку в точке R. Если диаметр окружности достаточно велик (обычно он равен по порядку величины 1 *м*), то, не делая заметной ошибки, можно допустить, что R лежит на окружности K. Следовательно, так как C — центр кривизны решетки, то угол падения SRC, а значит, и угол отражения снова равны α. Кроме того, дуга CP равняется дуге SC, и поэтому луч, отраженный в точке R, также проходит через P.

Аналогичные рассуждения применимы и к дифрагировавшим лучам. Пусть β — угол, который дифрагировавший в Q луч составляет с QP. Соответствующий луч того же порядка, дифрагировавший в R, составит такой же угол

(β) с RP. Таким образом, углы, образованные лучом, дифрагировавшим в Q, с SQ и лучом, дифрагировавшим в R, с SR, равны $2\alpha+\beta$. Следовательно, эти два луча приходят в точку P', лежащую на окружности K. Итак, *для получения резких линий щель, решетку и плоскость наблюдения (фотографическую пластинку) нужно располагать на окружности с диаметром, равным радиусу кривизны вогнутой решетки.*

На этом принципе основано несколько схем различных установок дифракционных решеток. Сам Роуланд использовал установку, схематически показан-

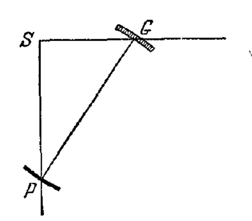


Рис. 8.27. Схема Роуланда для вогнутой решетки.

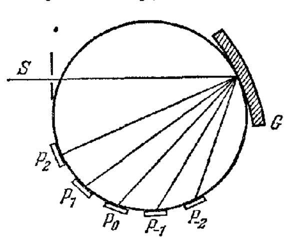


Рис. 8.28 Схема Пашена для вогнутой решетки

ную на рис. 8.27. Решетка G и кассета P с пластинкой укрепляются на противоположных концах подвижного стержня, длина которого равна радиусу кривизны решетки. Оба конца стержня свободно перемещаются вдоль неподвижных направляющих, расположенных под прямым углом друг к другу. Щель S устанавливается непосредственно над их пересечением таким образом, что свет, падающий нормально на щель, далее распространяется вдоль SG. Следовательно, щель располагается на круге Роуланда с диаметром PG и порядок спектров, образующихся на пластинке, зависит от положения стержня.

Другое расположение, показанное на рис. 8.28, позволяет устранить подвижные детали. Здесь роль круга Роуланда играет стальной рельс в виде кольца, на котором неподвижно укреплены щель S и решетка G. Вдоль кольца установлен ряд кассет P_0, P_1, P_{-1}, ..., что позволяет одновременно фотографировать спектры разных порядков. Такая установка, называемая *установкой Пашена*, отличается, кроме того, и большой стабильностью.

Схема, предложенная Иглом, напоминает по своей компактности, установку Литрова с плоской решеткой. Щель можно располагать выше или ниже центра кассеты с пластинкой (рис. 8.29) или даже сбоку кассеты, а свет можно направить в нужном направлении небольшим зеркальцем. Для наблюдения различных участков спектра пластинку и решетку нужно одновременно поворачивать на один и тот же угол, но в противоположных направлениях, а расстояние между ними следует изменять так, чтобы их поверхности всегда были касательны к кругу Роуланда. Наблюдается часть спектра, образованного лучами, дифрагировавшими обратно в направлении щели под углами, близкими к углу падения. Чтобы это устройство было строго автоколлимационным, щель S должна находиться в центре пластинки.

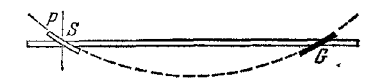


Рис. 8.29. Схема Игла для вогнутой решетки.

Изображения спектральных линий, полученные с вогнутой решеткой, обладают теми же аберрациями, что и изображения от вогнутого зеркала; среди них главная — астигматизм. Однако если вогнутая решетка освещается параллельным пучком света, то астигматизм можно уничтожить на нор-

мали к решетке и свести к очень малой величине по всему полезному спектру *).

8.6.2. Разрешающая сила систем, образующих изображение. Дифракционная формула Фраунгофера (8.3.36) находит важное применение при вычислении *разрешающей силы* оптических систем. Понятие разрешающей силы было введено нами в п. 7.6.3 в связи с интерференционной спектроскопией. Выше мы оценили разрешающую силу, которую можно достичь с решетками и призмами. Теперь мы распространим это понятие на оптические системы, образующие изображения.

Для спектральных аппаратов (например, для линейной решетки или интерферометра Фабри — Перо) разрешающая сила служит мерой способности прибора разделять две спектральные линии со слегка различающимися длинами волн. Для систем, образующих изображения, разрешающая сила является мерой способности прибора изображать раздельно две соседние точки объекта. Согласно законам геометрической оптики в отсутствие аберраций каждая точка объекта должна изображаться резкой точкой. Однако в результате дифракции она всегда будет иметь вид светлого пятнышка конечного размера. Если два таких пятнышка в изображении (дифракционные картины) начнут перекрываться, то чем ближе друг к другу центральные максимумы, тем труднее распознать наличие двух объектов. Предельное расстояние между двумя объектами, на котором глаз еще может обнаружить их существование, зависит в известной степени от опыта. Соответствующим проявлением фотографической пластинки можно увеличить контрастность изображения и тем самым увеличить предел разрешения. Тем не менее желательно иметь какой-нибудь простой критерий, позволяющий хотя бы грубо сравнивать относительную эффективность различных оптических систем. Воспользуемся снова критерием Рэлея, о котором говорилось в п. 7.6.3. Согласно этому критерию два изображения начнут разрешаться, если главный максимум одного совпадет с первым минимумом другого. Для спектральных аппаратов, для которых пределом разрешения служит некая разность длин волн $\Delta\lambda$, разрешающая сила определялась как величина $\lambda/\Delta\lambda$. Для систем, образующих изображения, пределом разрешения является некоторое расстояние δx или некоторый угол $\delta\theta$, а разрешающая сила определяется их обратной величиной (т. е. $1/\delta x$ или $1/\delta\theta$).

Рассмотрим сначала предел разрешения телескопа. Для удаленного объекта граница входного зрачка, совпадающая с круглой оправой объектива, действует как дифракционное отверстие. Если a — радиус апертуры объектива, то, согласно (8.5.16), положение первого минимума интенсивности относительно центрального максимума задается соотношением **)

$$w = 0{,}61\lambda/a. \qquad (24)$$

Здесь $w=\sqrt{p^2+q^2}$ — синус угла φ между направлением (p, q) и центральным направлением $p=q=0$. Обычно этот угол так мал, что его синус можно заменить самим углом, и тогда (на основании критерия Рэлея) получим, что *угловое расстояние двух звезд, которые начинают разрешаться, равно* $0{,}61\lambda/a$.

При данном объективе угловой размер изображения, видимого глазом, зависит от увеличения окуляра. Однако, применяя окуляр с бо́льшим увеличением, нельзя выявить детали, которых нет в первичном изображении; в самом деле, каждый элемент такого изображения представляет собой небольшую дифракционную картину, а изображение, видимое в окуляр, является только увеличенным изображением совокупности этих картин.

*) Такая свободная от астигматизма установка описана Водсвортом [52]. Теория аберрации решетки и установок с решетками рассматривается в [53].

**) Согласно п. 7.6.3 отношение интенсивностей в седловине и в максимуме на пределе разрешения при дифракции на щели равно $8/\pi^2 = 0{,}811$. Соответствующая величина в нашем случае (круглое отверстие) составляет 0,735.

Наибольший из существующих телескопов (в обсерватории Маунт Паломар) имеет зеркало диаметром $2a \approx 5$ *м*. Если пренебречь небольшим экранированием в центре телескопа, то для света с длиной волны, соответствующей центральной части видимой области ($\lambda = 5{,}6 \cdot 10^{-5}$ *см*), получим предел разрешения, равный

$$\varphi \approx 0{,}61 \frac{5{,}6 \cdot 10^{-5} \text{ см}}{2{,}5 \cdot 10^2 \text{ см}} \approx 1{,}4 \cdot 10^{-7},$$

или угловое разрешение (в секундах)

$$\varphi \approx 0{,}028''.$$

В § 6.1 в качестве предельного разрешения глаза была принята величина в 1′. Теперь можно оценить эту величину более точно. В зависимости от интенсивности света диаметр зрачка глаза изменяется примерно от 1,5 до 6 *мм*. Следовательно (снова принимая $\lambda = 5{,}6 \cdot 10^{-5}$ *см*), предел разрешения глаза находится в пределах

$$0{,}61 \frac{5{,}6 \cdot 10^{-5}}{0{,}75 \cdot 10^{-1}} > \varphi > 0{,}61 \frac{5{,}6 \cdot 10^{-5}}{3 \cdot 10^{-1}},$$

т. е.

$$4{,}55 \cdot 10^{-4} > \varphi > 1{,}14 \cdot 10^{-4},$$

или (в минутах и секундах)

$$1'34'' > \varphi > 0'24''.$$

До сих пор мы считали, что дифракционное отверстие имеет круглую форму. Значительный интерес представляет также кольцевое отверстие, так как во многих телескопах центральная часть круглого отверстия часто закрыта вторичным зеркалом. Предположим, что кольцевое отверстие ограничено двумя концентрическими окружностями с радиусами a и εa, где ε — некоторое положительное число, меньшее единицы. Распределение света в дифракционной картине Фраунгофера определяется интегралом вида (8.5.8), где интегрирование по ρ производится только по области $\varepsilon a \leqslant \rho \leqslant a$. Тогда вместо (8.5.13) получим

$$U(P) = C\pi a^2 \left[\frac{2J_1(kaw)}{kaw}\right] - C\pi\varepsilon^2 a^2 \left[\frac{2J_1(k\varepsilon aw)}{k\varepsilon aw}\right], \tag{25}$$

а интенсивность задается соотношением

$$I(P) = \frac{1}{(1-\varepsilon^2)^2} \left[\left(\frac{2J_1(kaw)}{kaw}\right) - \varepsilon^2 \left(\frac{2J_1(k\varepsilon aw)}{k\varepsilon aw}\right)\right]^2 I_0, \tag{26}$$

где $I_0 = |C|^2 \pi^2 a^4 (1-\varepsilon^2)^2$ — интенсивность в центре $w = 0$ картины. Положение минимумов (нулей) интенсивности здесь определяется корнями уравнения

$$J_1(kaw) - \varepsilon J_1(k\varepsilon aw) = 0 \qquad (w \neq 0), \tag{27}$$

тогда как положение максимумов — корнями уравнения

$$J_2(kaw) - \varepsilon^2 J_2(k\varepsilon aw) = 0. \tag{28}$$

При выводе последнего соотношения использовалось (8.5.15), как и в случае круглого отверстия. Для неэкранированного отверстия ($\varepsilon = 0$) первый корень (27), конечно, равен величине, определяемой выражением (24), а именно $w = 3{,}83/ka = 0{,}61\,\lambda/a$. При увеличении ε первый корень (27) уменьшается*); так, например, если $\varepsilon = 1/2$, то он становится немного меньше $3{,}15/ka = 0{,}50\,\lambda/a$. Так как главный максимум независимо от ε остается при $w = 0$, то очевидно, что при загораживании центральной части отверстия разрешающая сила

*) С помощью простых вычислений находим из (26), что если $\varepsilon \to 1$, то $I/I_0 \to J_0^2(kaw)$. Так как первый нуль уравнения $J_0(x) = 0$ получается при $x = 2{,}40$, то, следовательно, с увеличением ε радиус первого темного кольца приближается к величине, задаваемой $w = 2{,}40/ka = 0{,}38\,\lambda/a$.

возрастает. Ее увеличение сопровождается, однако, уменьшением яркости изображения; кроме того, в этом случае вторичные максимумы становятся более отчетливыми, что ведет к уменьшению контрастности. При $\varepsilon = 1/2$ первый вторичный максимум (при $w = 4{,}8/ka$) равен 0,10 главного максимума, тогда как для круглого отверстия он равен 0,018 (при $w = 5{,}14/ka$) (рис. 8.30).

Экранирование центральной части круглого отверстия соответствует замене функции зрачка (см. (8.3.39)) $G(\xi, \eta)$, равной C или 0 в соответствии с $0 \leqslant \rho \leqslant a$ или $\rho > a$, на $G(\xi, \eta)$, равной C или 0 в соответствии с $\varepsilon a \leqslant \rho \leqslant a$ или $\rho < \varepsilon a$, $\rho > a$. Конечно, можно изменить функцию зрачка и другим путем. Общий метод ее изменения состоит в том, что на одну или несколько поверхностей системы наносится тонкая полупрозрачная пленка подходящего вещества. Такого же эффекта можно достичь с помощью специально сконструированных «фильтров», например полой линзы соответствующей формы, наполненной поглощающей жидкостью. Здесь возникает задача определения вида функции зрачка, которая в некотором условном смысле давала бы наилучшее возможное изображение. Этот вопрос исследовался рядом ученых (см., например, [55]; хороший обзор исследований в этой области содержится в статье Вольфа [56]). Особый интерес представляет работа [57], где показано, что функцию зрачка можно выбрать так, чтобы радиус первого темного кольца был как угодно мал и в то же время темная зона, окружающая центральное кольцо, была бы сколь угодно велика. Однако постепенное уменьшение радиуса первого темного кольца приводит к постепенному уменьшению яркости центрального диска; отсюда следует, что наименьшие возможные размеры, а следовательно, и разрешающая сила ограничены количеством поступающего света.

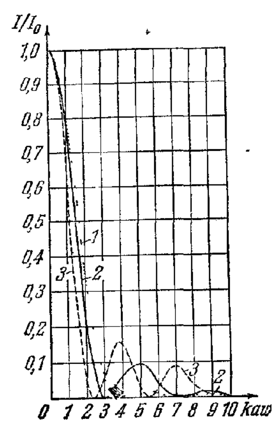


Рис. 8.30. Влияние диафрагмирования центральной части отверстия на разрешающую силу (по [54])

Нормированные кривые интенсивности в дифракционной картине Фраунгофера 1 — круглое отверстие, 2 — кольцевое отверстие с $\varepsilon = 1/2$, 3 — кольцевое отверстие с $\varepsilon \to 1$

Частичное подавление вторичных максимумов разумно выбранной функцией зрачка называется *аподизацией* *). В спектроскопии это облегчает обнаружение сателлитов спектральных линий, а в астрономии — разрешение двойных звезд с сильно различающейся видимой яркостью.

Принятая теория разрешающей силы, кратко изложенная в настоящем разделе, особенно применима к прямым визуальным наблюдениям. При других методах наблюдения (например, при фотометрическом методе) часто удается обнаружить существование двух объектов с угловым расстоянием, значительно меньшим указанного критерием Рэлея. В связи с этим интересно также сравнить разрешающие силы телескопа и звездного интерферометра Майкельсона (см. п. 7.3.6). Если о существовании двух звезд судят по первому исчезновению полос, образованных в интерферометре, и если максимальное расстояние между внешними зеркалами последнего равно d, то, согласно (7.3.42), пользуясь таким прибором, можно обнаружить двойные звезды с угловым расстоянием между ними вплоть до $\varphi \approx \frac{1}{2}\lambda/d$. Сравнение найденной величины с (24) показывает, что для обнаружения двойных звезд с таким разделением при визуальном наблюдении в телескоп необходимо, чтобы диаметр его объектива $2a$ примерно равнялся $2{,}4d$.

*) Полный обзор такого рода исследований дан в [58].

8.6.3. Образование изображения в микроскопе*). В нашем элементарном изложении теории разрешающей силы предполагалось, что свет от двух точек предмета некогерентен. Это предположение справедливо, если оба объекта самосветящиеся, как, например, звезды, наблюдаемые в телескоп. Получающаяся интенсивность в любой точке плоскости изображения равна тогда сумме интенсивностей, обусловленных каждой точкой предмета.

Как правило, при работе с микроскопом ситуация оказывается значительно сложнее. В большинстве случаев рассматриваемый предмет не является самосветящимся и, следовательно, должен освещаться с помощью вспомогательного устройства. Вследствие дифракции на отверстии осветительной системы (конденсора) каждый элемент источника создаст в предметной плоскости микроскопа дифракционную картину. Дифракционные картины с центрами в достаточно близких друг к другу точках частично перекрываются и, следовательно, в соседних точках плоскости предмета световые колебания в общем случае частично коррелированы. Часть этого света проходит сквозь предмет с изменением фазы или без него, тогда как оставшаяся его часть рассеивается, отражается или поглощается. Поэтому, вообще говоря, невозможно посредством одного наблюдения или, даже используя одно какое-то устройство, получить правильное увеличенное изображение всей микроструктуры объекта. По этой причине были разработаны различные методы наблюдения, пригодные для изучения определенных типов объектов или для выявления у них тех или иных характерных особенностей.

Изложим коротко теорию образования изображения в микроскопе, ограничиваясь двумя крайними случаями, а именно полностью некогерентным освещением и идеально когерентным освещением. Освещение частично когерентным светом рассматривается в п. 10.5.2.

а. Некогерентное освещение. Рассмотрим сначала самосветящийся объект, например накаленную нить электрической лампочки. Пусть P — осевая точка объекта, а Q — соседняя точка в плоскости предмета, находящаяся на расстоянии Y от P, и пусть P' и Q' — изображения этих точек (рис. 8.31). Далее пусть θ и θ' — углы между крайними лучами осевых пучков и осью.

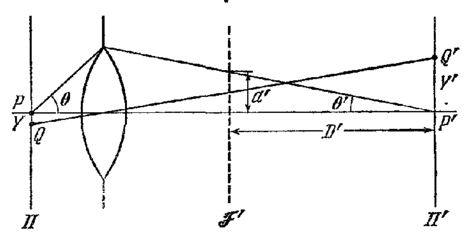


Рис. 8.31. К теории разрешающей силы микроскопа.

Если a' — радиус площадки (предполагается, что она круглая), которую образует сходящийся в P' пучок лучей при пересечении с задней фокальной плоскостью $\mathcal{F}'$, а D' — расстояние от выходного зрачка до плоскости изображения, то, поскольку θ' мало, можно написать

$$\theta' = a'/D'. \tag{29}$$

Далее, если $w = \sqrt{p^2 + q^2}$ — расстояние от Q' до P', выраженное в «дифракционных единицах» (см. (8.3.35) и (8.5.7)), т. е. это синус угла, под которым из центра дифракционного отверстия видна прямая, соединяющая эти точки, то с хорошим приближением мы можем написать

$$Y' = wD'. \tag{30}$$

Пусть n и n' — показатели преломления, а λ и λ' — длины волн в пространстве предмета и в пространстве изображения, а λ_0 — длина волны в вакууме. Тогда,

) Подробное изложение теории образования изображения в микроскопе дано в работах Л. И. Мандельштама и Д. С. Рождественского [119]. *(Прим. ред.)*

согласно (8.5.16), первый минимум дифракционной картины от P определяется величиной $w = 0{,}61\lambda'/a'$, и мы получим на пределе разрешения

$$Y' = 0{,}61\lambda' \frac{D'}{a'} = 0{,}61 \frac{\lambda'}{\theta'} = 0{,}61 \frac{\lambda_0}{n'\theta'}. \tag{31}$$

Само собою разумеется, микроскоп рассчитывается так, чтобы он давал резкое изображение не только какой-либо осевой точки, но и соседних с ней точек, лежащих в плоскости предмета. Поэтому, согласно п. 4.5.1, должно удовлетворяться условие синусов *)

$$nY \sin\theta = -n'Y' \sin\theta'.$$

Так как угол θ' мал, можно заменить $\sin\theta'$ на θ' и, подставляя найденное отсюда значение Y' в (31), получим окончательно

$$|Y| \approx 0{,}61\lambda_0/n \sin\theta. \tag{32}$$

Эта формула дает расстояние между двумя точками объекта, которые при освещении *некогерентным* светом начинают разрешаться микроскопом с круглым отверстием объектива.

Величина $n \sin\theta$, входящая в (32), называется *числовой апертурой* (см. (4.8.13)). Она должна быть большой, если необходима большая разрешающая сила. Пути достижения большой числовой апертуры рассматривались в § 6.6.

б. *Когерентное освещение*; *теория Аббе*. Ниже рассматривается другой крайний случай, а именно случай, когда свет, идущий от объекта, можно считать строго когерентным. Это положение приблизительно осуществляется при освещении тонкого предмета со сравнительно простой структурой светом небольшого источника, проходящим через конденсор с малой апертурой (см. ниже, п. 10.5.2).

Первая удовлетворительная теория разрешения при когерентном освещении была сформулирована Аббе ([59, 60]); хорошее изложение теории Аббе дано в [61]. Ему же принадлежат и прекрасные опыты, наглядно подтверждающие эту теорию. Согласно Аббе, предмет ведет себя как дифракционная решетка, и поэтому при определении комплексного возмущения в любых точках плоскости изображения должны учитываться не только все элементы отверстия объектива, но и все элементы самого предмета. Выражаясь математическим языком, можно сказать, что переход от предмета к изображению совершается с помощью двойного интегрирования: одного по предметной плоскости и другого по площади отверстия объектива. В теории Аббе в первую очередь рассматривается дифракция на предмете, а влияние апертуры учитывается во вторую очередь. Возможен также и обратный порядок, приводящий, естественно, к таким же результатам **).

Для иллюстрации теории Аббе рассмотрим вначале образование изображения предмета в виде решетки, на который перпендикулярно его плоскости падает плоская волна (центральное освещение по Кёлеру). В результате дифракции волны на такой решетке (см. п. 8.6.1) на задней фокальной плоскости $\mathcal{F}'$ объектива возникает дифракционная картина Фраунгофера. На рис 8 32 максимумы спектров последовательных порядков этой картины обозначены $\ldots S_{-2}, S_{-1}, S_0, S_1, S_2, \ldots$ Каждую точку в фокальной плоскости можно рассматривать как центр вторичного когерентного возмущения, величина которого пропорциональна амплитуде в этой точке. Световые волны, идущие от таких вторичных источников, интерферируя между собой, образуют изображение предмета в плоскости изображения Π' объектива. Для получения точного изображения предмета необходимо, чтобы все спектры участвовали в формирова-

*) Знак минус появился потому, что θ' соответствует величине $-\gamma_1$ в § 4 5.

**) Теория формирования изображения в микроскопе, эквивалентная такому подходу, была сформулирована Рэлеем в [62, 63].

нии изображения. Строго говоря, это невозможно, так как отверстие объектива всегда имеет ограниченный размер. Ниже мы увидим, что исключение некоторых спектров может привести к искажению вида изображения. Для практических целей, очевидно, совершенно достаточно отверстия такого размера, чтобы оно пропустило все спектры, несущие заметное количество энергии.

Представим эти соображения в более точном виде и не будем более ограничиваться предметом в виде решетки. Если x и y — координаты произвольной точки в плоскости предмета, f — расстояние между фокальной плоскостью $\mathcal{F}'$ и объективом, то возмущение в точке с координатами

$$\xi = pf, \qquad \eta = qf, \tag{33}$$

лежащей в плоскости $\mathcal{F}'$ (см. рис. 8.32), определяется формулой Фраунгофера

$$U(\xi, \eta) = C_1 \iint\limits_{\mathcal{A}} F(x, y) \exp\left\{-ik\left[\frac{\xi}{f}x + \frac{\eta}{f}y\right]\right\} dx\,dy, \tag{34}$$

где F — функция пропускания предмета, C_1 — постоянная, и интегрирование производится по площади $\mathcal{A}$, занятой предметом в плоскости предмета П.

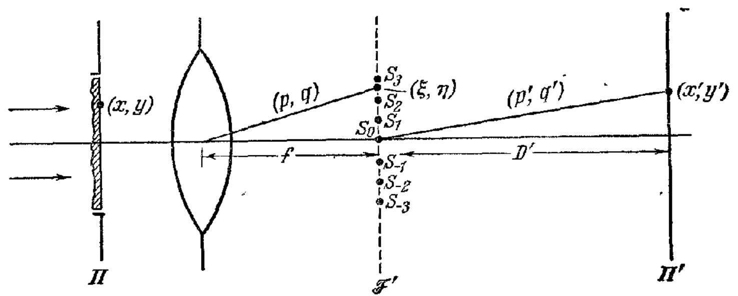


Рис. 8.32. К теории Аббе образования изображения в микроскопе при освещении когерентным светом.

Рассмотрим теперь переход от задней фокальной плоскости $\mathcal{F}'$ к плоскости изображений П'. Обозначим, как и прежде, через D' расстояние от $\mathcal{F}'$ до П', а через $V(x', y')$ — возмущение в произвольной точке

$$x' = p'D', \qquad y' = q'D' \tag{35}$$

плоскости изображений. Тогда для дифракции Фраунгофера на отверстии $\mathcal{B}$ в плоскости $\mathcal{F}'$ имеем, предположив, что $a'/D' \ll 1$ (см. рис. 8.31),

$$V(x', y') = C_2 \iint\limits_{\mathcal{B}} U(\xi, \eta) \exp\left\{-ik\left[\frac{x'}{D'}\xi + \frac{y'}{D'}\eta\right]\right\} d\xi\,d\eta. \tag{36}$$

Подставляя сюда (34), получим

$$V(x', y') = C_1C_2 \iint\limits_{\mathcal{A}} \iint\limits_{\mathcal{B}} F(x, y) \exp\left\{-i\frac{k}{f}\left[\left(x + \frac{f}{D'}x'\right)\xi + \left(y + \frac{f}{D'}y'\right)\eta\right]\right\} dx\,dy\,d\xi\,d\eta. \tag{37}$$

Итак, если $F(x, y)$ считается равной нулю во всех точках плоскости предмета, лежащих вне $\mathcal{A}$, то интегрирование по x и y можно формально производить от $-\infty$ до $+\infty$. Подобным же образом, если отверстие $\mathcal{B}$ так велико, что $|U(\xi, \eta)|$ пренебрежимо мало для точек плоскости $\mathcal{F}'$, лежащих вне $\mathcal{B}$, то интегрирование по ξ и η также можно производить от $-\infty$ до $+\infty$. Отметим, кроме того (см.

(4.3.10), где f' и Z' соответствуют нашим $-f$ и D'), что

$$\frac{f}{D'} = -\frac{1}{M}, \tag{38}$$

где M (<0) увеличение П′ относительно П, применяя интегральную теорему Фурье (см., например, [64]), получим

$$V(x', y') = CF\left(\frac{x'}{M}, \frac{y'}{M}\right) = CF(x, y), \tag{39}$$

где (x, y) — точка предмета, изображение которой находится в (x', y'), а

$$C = C_1 C_2 \lambda^2 f^2$$

— постоянная. Следовательно, с принятой здесь точностью *) изображение будет полностью подобно предмету (но перевернуто), если апертура достаточно велика.

Нетрудно показать, что в изображении могут появляться совершенно искаженные детали, если исключены некоторые спектры, несущие заметную долю энергии. Для этого рассмотрим предмет в виде решетки с периодом d, состоящей из N эквидистантных конгруэнтных щелей шириной s, разделенных непрозрачными промежутками. Для простоты предполагается, что диафрагма имеет прямоугольную форму, и две ее стороны параллельны щелям.

Согласно (8.6.3) находим

$$U(\xi) = C_1\left(\frac{\sin\frac{k\xi s}{2f}}{\frac{k\xi s}{2f}}\right)\frac{1-\exp(-iNk\,d\xi/f)}{1-\exp(-ik\,d\xi/f)}, \tag{40}$$

где $U^{(0)}$ заменено выражением, относящимся к дифракции на прямоугольном отверстии (см. п. 8.5.1). Если прямоугольное отверстие занимает в направлении ξ область

$$-a \leqslant \xi \leqslant a,$$

то в соответствии с (6) и (40) возмущение в плоскости изображений равно

$$V(x') = C_1 C_2 \int_{-a}^{a} \frac{\sin\frac{ks\xi}{2f}}{\frac{ks\xi}{2f}}\,\frac{1-\exp(-iNk\,d\xi/f)}{1-\exp(-ik\,d\xi/f)}\exp(-ikx'\xi/D')\,d\xi. \tag{41}$$

Положение главных максимумов интенсивности определяется корнями уравнения $1-\exp[-ik\,d\xi/f]=0$, т. е. $\xi = mf\lambda/d$, где m — целое число. Между этими главными максимумами находятся слабые вторичные максимумы. При большом N главные максимумы очень остры и по сравнению с ними вторичные максимумы пренебрежимо малы. В таком случае с хорошим приближением можно заменить этот интеграл суммой интегралов, каждый из которых берется от средней точки Q_m интервала между соседними главными максимумами до средней точки Q_{m+1} следующего такого же интервала. В каждом интервале мы можем заменить аргумент его значением в центре $\xi = mf\lambda/d = 2\pi mf/kd$,

$$V(x') \approx V_0 \sum_{-\bar{m}<m<\bar{m}} \frac{\sin\frac{m\pi s}{d}}{\frac{m\pi s}{d}} \exp\left\{\frac{2\pi i m x'}{d'}\right\}. \tag{42}$$

*) Как отмечалось выше, использованное здесь приближение Фраунгофера ограничивается случаем, когда точки предмета и точки изображения лежат достаточно близко к оси.

Здесь

$$\bar{m}=\frac{ad}{\lambda f}, \quad d'=Md=-\frac{D'}{f}d, \tag{43}$$

а V_0 — интеграл

$$V_0=C_1C_2\int\limits_{Q_m}^{Q_{m+1}}\frac{1-\exp(-iNk\,d\xi/f)}{1-\exp(-ik\,d\xi/f)}\,d\xi, \tag{44}$$

который, не считая небольшого поправочного члена в высоких порядках, практически не зависит от m. Ряд (42) можно переписать в действительной форме, а именно

$$\frac{V(x')}{V_0}=1+2\sum_{1<m<\bar{m}}\frac{\sin\frac{m\pi s}{d}}{\frac{m\pi s}{d}}\cos\frac{2\pi mx'}{d'}. \tag{45}$$

Предположим сперва, что длина a диафрагмы очень велика. Тогда суммирование формально можно распространить на весь бесконечный ряд ($\bar{m}=\infty$);

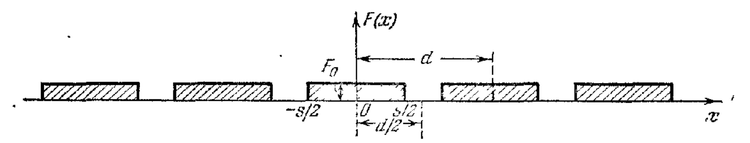


Рис. 8.33. Объект типа решетки.

легко доказать, что при таких условиях изображение точно подобно предмету. Для этого разложим функцию пропускания F предмета типа решетки (рис. 8.33)

$$\left.\begin{aligned} F(x)&=F_0, &\text{когда}\quad 0<|x|<\frac{s}{2},\\ F(x)&=0, &\text{когда}\quad \frac{s}{2}<|x|<\frac{d}{2} \end{aligned}\right\} \tag{46}$$

в ряд Фурье

$$F(x)=c_0+2\sum_{m=1}^{\infty}c_m\cos\frac{2\pi mx}{d}; \tag{47}$$

тогда

$$c_0=\frac{F_0s}{d}, \quad c_m=F_0\frac{\sin\frac{\pi ms}{d}}{\pi m} \quad (m=1,\ 2,\ 3,\ \ldots). \tag{48}$$

С точностью до постоянного множителя этот ряд не отличается от (45).

Предположим теперь, что длина a диафрагмы уменьшается. Если a так мало, что в образовании изображения участвует только спектр нулевого порядка, т. е. если $\bar{m}=ad/\lambda f$ — правильная дробь, то, согласно (45), $V(x')=\text{const}$, и плоскость изображений освещена равномерно. (Последний вывод, конечно, не совсем правилен, так как мы отбросили некоторые остаточные члены; на самом же деле наблюдается небольшое падение интенсивности в направлении края изображения.)

Если, кроме спектра нулевого порядка, сквозь диафрагмы проходят еще два спектра первого порядка (S_1, S_{-1}), т. е. если $\bar{m}=ad/\lambda f$ немного больше

единицы, то из (45) находим

$$\frac{V(x')}{V_0}=1+2\frac{\sin\frac{\pi s}{d}}{\frac{\pi s}{d}}\cos\frac{2\pi x'}{d'}. \tag{49}$$

Теперь изображение имеет правильную периодичность $x'=d'$, но значительно сглаженное распределение интенсивности. При увеличении диафрагмы изображение становится все более и более похожим на предмет.

Совершенно искаженное изображение наблюдается в том случае, когда в его образовании спектры низших порядков совсем не участвуют. Если, например, исключены все порядки, кроме второго, то

$$\frac{V(x')}{V_0}=2\frac{\sin\frac{2\pi s}{d}}{\frac{2\pi s}{d}}\cos\frac{4\pi x'}{d'}, \tag{50}$$

и изображение имеет период $x'=d'/2$, т. е. в таком «изображении» наблюдается вдвое больше линий, чем действительно есть у предмета.

Определим теперь разрешающую силу микроскопа. Вернемся опять к рассмотрению рис. 8.31, но допустим, что свет, идущий из P и Q, когерентен. В этом случае распределение интенсивности в плоскости изображений определяется в основном когерентной суперпозицией двух дифракционных картин Эйри с центром одной в P', а другой в Q'. Комплексная амплитуда в точке, находящейся между P' и Q' на расстоянии w_1 (измеренном в «дифракционных единицах») от P' определяется соотношением

$$U(w_1)=\left[\frac{2J_1(kaw_1)}{kaw_1}+\frac{2J_1[ka(w-w_1)]}{ka(w-w_1)}\right]U_0, \tag{51}$$

где w — расстояние между P' и Q'; другие обозначения имеют такой же смысл, как и раньше. Следовательно, для интенсивности имеем

$$I(w_1)=\left[\frac{2J_1(kaw_1)}{kaw_1}+\frac{2J_1(ka(w-w_1))}{ka(w-w_1)}\right]^2 I_0. \tag{52}$$

В случае некогерентного освещения P' и Q' считались разрешенными, если главный максимум интенсивности одной картины совпадал с первым минимумом другой. Интенсивность в средней точке ($w\approx 1,92$) между двумя максимумами равняется тогда $2[2J_1(1,92)/1,92]^2\approx 0,735$ от интенсивности каждого максимума, т. е. кривая суммарной интенсивности спадает на участке между двумя главными максимумами примерно на 26,5%. (Это соответствует 19% для отверстия в виде щели, см. рис. 7.62.) Если мы снова примем, что такое уменьшение интенсивности по существу определяет предел разрешения, то критическое расстояние $w=2w_1$ находится из соотношения

$$\frac{I(w_1)}{I_0}=\left[2\frac{2J_1(kaw_1)}{kaw_1}\right]^2=0,735. \tag{53}$$

Первый корень этого трансцендентного уравнения равен $w_1\approx 2,41/ka$, и, следовательно, для критического расстояния, измеренного в обычных единицах, имеем

$$Y'=2w_1D'\approx\frac{2,41}{\pi}\frac{D'}{a}=\frac{0,77\lambda'}{\theta'}=\frac{0,77\lambda_0}{n'\theta'}. \tag{54}$$

Для того чтобы перейти от Y' к соответствующему расстоянию Y между точками предмета, используем условие синусов (допуская, что $\sin\theta'\approx\theta'$) и получим окончательно, что *предел разрешения при когерентном освещении* равен

$$|Y|=0,77\frac{\lambda_0}{n\sin\theta}. \tag{55}$$

Не считая несколько большего численного множителя (который всегда до известной степени произволен, так как он зависит от формы предмета и отверстия и от чувствительности приемника), мы получим то же выражение, что и в случае некогерентного освещения (см. уравнение (32)). Таким образом, разрешающая сила для света определенной длины волны по существу зависит только от числовой апертуры объектива.

в. *Когерентное освещение; фазо-контрастный метод наблюдения Цернике* *). Мы определили *фазовый объект* как объект, который изменяет фазу, а не амплитуду падающего света. Объектом такого типа служит неоднородный по оптической толщине и вместе с тем совершенно не поглощающий света предмет. Такие объекты часто встречаются в биологии, кристаллографии и других областях науки. Из предыдущего очевидно, что обычные методы наблюдения дают мало информации о таких фазовых объектах. Комплексная амплитудная функция, определяющая возмущение в плоскости изображения, в этом случае подобна функции пропускания объекта **). Поскольку глаз (или другой регистрирующий прибор) способен отмечать только изменение интенсивности, можно делать выводы об изменении амплитуды, но не об изменениях фазы, вызванных объектом.

Для получения данных о фазовых объектах следует применять специальные методы наблюдения, например так называемый *метод темного поля*, где с помощью диафрагмы устраняется центральный порядок, или метод *свилей*, в котором исключаются все спектры по одну сторону от центрального порядка. Однако наиболее эффективным является метод, предложенный Цернике [67], впервые описанный им в 1935 г. и известный под названием метода *фазового контраста*. Его преимущество заключается в том, что распределение интенсивности в изображении прямо пропорционально изменению фазы, вносимому соответствующими точками объекта.

Чтобы объяснить принцип метода фазового контраста, рассмотрим сначала прозрачный объект в виде одномерной фазовой решетки. Функция пропускания такого объекта по определению (см. стр. 369) имеет вид

$$F(x) = \exp[i\varphi(x)], \tag{56}$$

где $\varphi(x)$ — вещественная периодическая функция с периодом (например, d), равным периоду решетки. Предположим, что величина φ мала по сравнению с единицей, и поэтому можно написать

$$F(x) \approx 1 + i\varphi(x). \tag{57}$$

Разлагая F в ряд Фурье

$$F(x) = \sum_{m=-\infty}^{\infty} c_m \exp\left(\frac{2\pi imx}{d}\right) \tag{58}$$

и учитывая, что F имеет вид (56), а φ — вещественная, но малая по сравнению с единицей величина, получим

$$c_0 = 1, \qquad c_{-m} = -c_m^* \qquad (m \neq 0). \tag{59}$$

Интенсивность спектров порядка m пропорциональна $|c_m|^2$.

В методе фазового контраста в заднюю фокальную плоскость $\mathcal{F}'$ объектива помещают тонкую пластинку из прозрачного вещества, называемую *фазовой пластинкой*. Такая пластинка вызывает отставание или опережение фазы в нулевом порядке (S_0 на рис. 8.32) относительно фазы в дифракционных спектрах

*) Более подробное изложение фазо-контрастного метода приведено, например, в [65].

**) Точное подобие достигается только при бесконечно большом отверстии объектива. Так как оно всегда конечно, можно увидеть некоторые детали фазовой структуры. Иногда при небольшой расфокусировке прибора видность таких «изображений» улучшается за счет разрешения (см. [66]).

(S_1, S_{-1}, S_2, S_{-2}, ...) на четверть периода. Это значит, что распределение комплексной амплитуды в фокальной плоскости будет отличаться от распределения, характеризующегося коэффициентами c_m, и станет распределением, характеризующимся коэффициентами c'_m, где

$$c'_0 = c_0 \exp\left[\pm\frac{i\pi}{2}\right] = \pm i, \qquad c'_m = c_m \qquad (m \neq 0). \tag{60}$$

Положительный или отрицательный знак берется в зависимости от того, получается ли отставание или опережение фазы в нулевом порядке. Теперь результирующее распределение света в плоскости изображений представляет не фазовую решетку (57), а фиктивную амплитудную решетку

$$G(x) = \pm i + i\varphi(x). \tag{61}$$

Следовательно, пренебрегая φ^2 по сравнению с единицей, находим, что интенсивность в плоскости изображения пропорциональна

$$I(x') = |G(x)|^2 = 1 \pm 2\varphi(x), \tag{62}$$

где, как и раньше, $x' = Mx$, а M — увеличение. Отсюда следует, что *в случае наблюдения методом фазового контраста, изменение фазы, вносимое объектом, превращается в изменение интенсивности, а последняя в любой точке плоскости изображений прямо пропорциональна (с точностью до аддитивной постоянной) изменению фазы, вызванному соответствующим элементом объекта* *). При отставании фазы в нулевом порядке относительно фазы дифракционных спектров (верхний знак в (61)) области объекта с большей оптической толщиной кажутся ярче мест со средней освещенностью. Это *светлый фазовый контраст*. Если фаза в нулевом порядке опережает фазы дифракционных спектров, области бо́льших оптических толщин кажутся темнее общего фона. Это так называемый *темный фазовый контраст* (рис. I и рис. II на вклейке).

Часто для получения хорошего разрешения отверстие осветительной системы делают кольцевым, а не круглым (см. п. 8.6.2). В таком случае роль спектра нулевого порядка S_0 на рис. 8.32 играет кольцевая область $\mathcal{T}'$, сквозь которую проходит прямой (не дифрагированный) свет, причем фаза именно этого света должна отставать или опережать на четверть периода фазы дифракционных спектров.

Фазовую пластинку можно изготовить, нанося испарением тонкий слой соответствующего диэлектрика на стеклянную подложку. Если свет, прошедший сквозь пластинку, должен отставать по фазе на четверть периода, то толщина нанесенного на пластинку слоя d должна составлять $d = \lambda/4(n-1)$, где n — показатель преломления диэлектрика. Отставание по фазе на четверть периода нулевого порядка, конечно, эквивалентно опережению по фазе дифракционных спектров на три четверти периода и наоборот. Чувствительность метода фазового контраста можно увеличить, если применять слабо поглощающие покрытия вместо диэлектрических. К этому вопросу мы еще вернемся позднее.

Остается показать, что применение метода фазового контраста не ограничивается только фазовыми объектами периодической структуры. Для этого разделим интеграл (34) на две части, а именно

$$U(\xi, \eta) = U_0(\xi, \eta) + U_1(\xi, \eta), \tag{63}$$

где

$$\left.\begin{aligned} U_0 &= C_1 \iint_{\mathcal{A}} \exp\left\{-\frac{ik}{f}[\xi x + \eta y]\right\} dx\,dy, \\ U_1 &= C_1 \iint_{\mathcal{A}} [F(x, y) - 1] \exp\left\{-\frac{ik}{f}[\xi x + \eta y]\right\} dx\,dy. \end{aligned}\right\} \tag{64}$$

*) Приближения, принятые в элементарной теории метода фазового контраста рассмотрены в [68, 69].

U_0 представляет распределение света в плоскости $\mathcal{F}'$ в отсутствие объекта, а U_1— влияние дифракции. «Прямой свет» U_0 (соответствующий нулевому порядку S_0 на рис. 8.32) концентрируется только в небольшой области $\mathcal{B}_0$ плоскости $\mathcal{F}'$ вблизи осевой точки $\xi=\eta=0$. Вместе с тем в общем случае только очень небольшая часть дифрагировавшего света попадает в эту область; большая его часть окажется в других местах *) этой плоскости $\mathcal{F}'$.

Предположим, что область $\mathcal{B}_0$, через которую проходит прямой свет, закрыта фазовой пластинкой. Действие последней описывается функцией пропускания

$$A=ae^{i\alpha}. \tag{65}$$

Для пластинки, вызывающей только изменение фазы проходящего света, $a=1$, а для пластинки, еще и поглощающей свет, $a<1$. Для света, прошедшего сквозь отверстие объектива, можно написать

$$U'(\xi,\ \eta)=AU_0(\xi,\ \eta)+U_1(\xi,\ \eta); \tag{66}$$

поэтому, согласно (36), распределение комплексной амплитуды в изображении имеет вид

$$V(x',\ y')=V_0(x',\ y')+V_1(x',\ y'), \tag{67}$$

где

$$\left.\begin{aligned} V_0&=AC_2\iint\limits_{\mathcal{B}} U_0(\xi,\ \eta)\exp\left\{-\frac{ik}{D'}[x'\xi+y'\eta]\right\}d\xi\,d\eta,\\ V_1&=C_2\iint\limits_{\mathcal{B}} U_1(\xi,\ \eta)\exp\left\{-\frac{ik}{D'}[x'\xi+y'\eta]\right\}d\xi\,d\eta. \end{aligned}\right\} \tag{68}$$

Так как отверстие $\mathcal{B}$ значительно превышает размеры участка $\mathcal{B}_0$, а U_0 вне $\mathcal{B}_0$ практически равно нулю, то, не вводя заметной ошибки, можно распространить область интегрирования в выражении для V_0 на всю плоскость $\mathcal{F}'$. Кроме того, если допустить, что $\mathcal{B}$ велико и пропускает все дифрагировавшие лучи, несущие заметную энергию, то у интеграла V_1 также можно написать бесконечные пределы. Наконец, если, как и раньше, функция пропускания $F(x,\ y)$ определена как нуль в точках плоскости предмета, лежащих вне области, занятой объектом, то интегралы (64) также можно взять с бесконечными пределами. Тогда, подставляя (64) в (68) и используя интегральную теорему Фурье и соотношение (38), получим

$$\left.\begin{aligned} V_0(x',\ y')&=CA,\\ V_1(x',\ y')&=C\left[F\left(\frac{x'}{M},\ \frac{y'}{M}\right)-1\right]=C\,[F(x,\ y)-1]. \end{aligned}\right\} \tag{69}$$

Из (67) и (69) следует, что интенсивность в плоскости изображений определяется выражением

$$I(x',\ y')=|V(x',\ y')|^2=|C|^2\,|A+F(x,\ y)-1|^2. \tag{70}$$

При наличии фазового объекта имеем

$$F(x,\ y)=\exp[i\varphi(x,\ y)], \tag{71}$$

и (70) сводится к **)

$$I(x',\ y')=|C|^2\,[a^2+2\{1-a\cos\alpha-\cos\varphi(x,\ y)+a\cos(\alpha-\varphi(x,\ y))\}]. \tag{72}$$

Так как мы предположили, что φ мало, последнее соотношение можно переписать в виде

$$I(x',\ y')=|C|^2\,[a^2+2a\varphi(x,\ y)\sin\alpha], \tag{73}$$

*) Этот вопрос подробно рассмотрен в [68]; см. также [72], где он исследовался в другой связи.

**) Особый случай, когда $a=0$, соответствует наблюдению в темном поле. Согласно (72) распределение интенсивности тогда определяется выражением

$$I(x',\ y')=2C^2\,[1-\cos\varphi(x,\ y)].$$

и если разность фаз, вносимая пластинкой, соответствует отставанию или опережению на четверть периода, то $\alpha = \pm\pi/2$, и (73) сводится к

$$I(x', y') = |C|^2 [a^2 \pm 2a\varphi(x, y)]. \tag{74}$$

Если пластинка не поглощает света ($a = 1$), мы снова получим выражение (62). В этом случае изменения интенсивности прямо пропорциональны изменению фаз, вносимому объектом. Для пластинки, поглощающей долю всего прямого света, равную a^2, отношение второго члена к первому в (73) равно $\pm\varphi/a$, что указывает на увеличение контрастности изображения. Например, ослабление прямого света в девять раз повышает чувствительность этого метода в три раза.

§ 8.7. Дифракция Френеля на прямолинейном крае

8.7.1. Дифракционный интеграл. После того как мы рассмотрели различные случаи дифракции Фраунгофера, займемся более общим случаем, а именно дифракцией Френеля.

Основной дифракционный интеграл (8.3.28) можно записать в виде

$$U(P) = B(C + iS), \tag{1}$$

где

$$B = -A\frac{i}{\lambda}\cos\delta\,\frac{\exp[ik(r'+s')]}{r's'}, \tag{2}$$

$$C = \iint_{\mathcal{A}} \cos\{kf(\xi, \eta)\}\,d\xi\,d\eta, \qquad S = \iint_{\mathcal{A}} \sin\{kf(\xi, \eta)\}\,d\xi\,d\eta. \tag{3}$$

Тогда интенсивность $I(P) = |U(P)|^2$ в точке наблюдения P равняется

$$I(P) = B^2(C^2 + S^2). \tag{4}$$

Теперь в разложении $f(\xi, \eta)$ (8.3.31) мы должны сохранить члены с ξ и η по крайней мере до второго порядка.

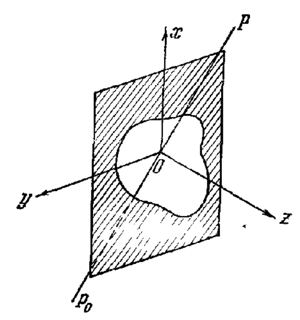


Рис. 8.34. К дифракции Френеля на отверстии в плоском непрозрачном экране.

Как и раньше, мы считаем, что плоскость отверстия $\mathcal{A}$ совпадает с плоскостью xy. Для упрощения расчетов направим ось x вдоль проекции линии P_0P на плоскость отверстия (рис. 8.34). Следовательно, при заданном положении источника система координат, вообще говоря, будет различной для различных точек наблюдения.

Согласно (8.3.30) $l = l_0$, $m = m_0$, и линейные члены в $f(\xi, \eta)$ исчезают. Направляющие косинусы лучей P_0O и OP равны

$$l = l_0 = \sin\delta, \qquad m = m_0 = 0, \qquad n = n_0 = \cos\delta, \tag{5}$$

где, как и прежде, δ — угол между линией P_0P и осью z. Выражение (8.3.31) для $f(\xi, \eta)$ сводится к

$$f(\xi, \eta) = \frac{1}{2}\left(\frac{1}{r'} + \frac{1}{s'}\right)(\xi^2\cos^2\delta + \eta^2) + \dots \tag{6}$$

Отбрасывая члены с ξ и η в третьей и более высоких степенях, мы приведем интегралы (3) к виду

$$\left.\begin{aligned} C &= \iint_{\mathcal{A}} \cos\left\{\frac{\pi}{\lambda}\left(\frac{1}{r'} + \frac{1}{s'}\right)(\xi^2\cos^2\delta + \eta^2)\right\} d\xi\,d\eta, \\ S &= \iint_{\mathcal{A}} \sin\left\{\frac{\pi}{\lambda}\left(\frac{1}{r'} + \frac{1}{s'}\right)(\xi^2\cos^2\delta + \eta^2)\right\} d\xi\,d\eta. \end{aligned}\right\} \tag{7}$$

Для удобства введем новые переменные интегрирования u и v, определяемые соотношениями

$$\frac{\pi}{\lambda}\left(\frac{1}{r'}+\frac{1}{s'}\right)\xi^2\cos^2\delta=\frac{\pi}{2}u^2, \qquad \frac{\pi}{\lambda}\left(\frac{1}{r'}+\frac{1}{s'}\right)\eta^2=\frac{\pi}{2}v^2. \tag{8}$$

Тогда

$$d\xi\, d\eta=\frac{\lambda}{2}\frac{du\, dv}{\left(\frac{1}{r'}+\frac{1}{s'}\right)\cos\delta},$$

и наши интегралы принимают вид

$$C=b\iint_{\mathcal{A}'}\cos\left\{\frac{\pi}{2}(u^2+v^2)\right\}du\, dv, \qquad S=b\iint_{\mathcal{A}'}\sin\left\{\frac{\pi}{2}(u^2+v^2)\right\}du\, dv, \tag{9}$$

где

$$b=\frac{\lambda}{2\left(\frac{1}{r'}+\frac{1}{s'}\right)\cos\delta}. \tag{10}$$

Здесь интегрирование производится по области $\mathcal{A}'$ плоскости (u, v), в которую подстановкой (8) преобразована область $\mathcal{A}$ отверстия.

8.7.2. Интегралы Френеля. Пусть $\mathcal{A}'$ — прямоугольник со сторонами, параллельными осям u и v; тогда интегралы (9) можно упростить еще больше, используя следующие тождества:

$$\left.\begin{aligned}\cos\left[\frac{\pi}{2}(u^2+v^2)\right]&=\cos\left(\frac{\pi}{2}u^2\right)\cos\left(\frac{\pi}{2}v^2\right)-\sin\left(\frac{\pi}{2}u^2\right)\sin\left(\frac{\pi}{2}v^2\right),\\ \sin\left[\frac{\pi}{2}(u^2+v^2)\right]&=\sin\left(\frac{\pi}{2}u^2\right)\cos\left(\frac{\pi}{2}v^2\right)+\cos\left(\frac{\pi}{2}u^2\right)\sin\left(\frac{\pi}{2}v^2\right).\end{aligned}\right\} \tag{11}$$

В этом случае для вычисления (9) мы должны рассмотреть интегралы

$$\mathscr{C}(w)=\int_0^w\cos\left(\frac{\pi}{2}\tau^2\right)d\tau, \qquad \mathscr{S}(w)=\int_0^w\sin\left(\frac{\pi}{2}\tau^2\right)d\tau. \tag{12}$$

Интегралы $\mathscr{C}(w)$ и $\mathscr{S}(w)$ называются *интегралами Френеля*. Они имеют большое значение при решении многих дифракционных задач и хорошо изучены. Мы должны кратко познакомиться с некоторыми их особенностями *).

Сначала получим выражения для $\mathscr{C}(w)$ и $\mathscr{S}(w)$ в виде рядов. Для этого разложим косинус и синус под знаком интеграла в степенные ряды и, интегрируя почленно, найдем

$$\left.\begin{aligned}\mathscr{C}(w)&=w\left[1-\frac{1}{2!\,5}\left(\frac{\pi}{2}w^2\right)^2+\frac{1}{4!\,9}\left(\frac{\pi}{2}w^2\right)^4-\dots\right],\\ \mathscr{S}(w)&=w\left[\frac{1}{1!\,3}\left(\frac{\pi}{2}w^2\right)-\frac{1}{3!\,7}\left(\frac{\pi}{2}w^2\right)^3+\frac{1}{5!\,11}\left(\frac{\pi}{2}w^2\right)^5-\dots\right].\end{aligned}\right\} \tag{13}$$

Эти ряды сходятся при всех значениях w, однако для вычислений они пригодны только при малом w. Когда w велико, интегралы можно вычислить, пользуясь разложением в ряды по отрицательным степеням w. Перепишем (12) в виде

$$\mathscr{C}(w)=\mathscr{C}(\infty)-\int_w^\infty\frac{d}{d\tau}\left(\sin\frac{\pi}{2}\tau^2\right)\frac{d\tau}{\pi\tau}. \tag{14}$$

Интегрируя по частям, получим

$$\mathscr{C}(w)=\mathscr{C}(\infty)+\frac{1}{\pi w}\sin\left(\frac{\pi}{2}w^2\right)+\int_w^\infty\frac{d}{d\tau}\left(\cos\frac{\pi}{2}\tau^2\right)\frac{d\tau}{\pi^2\tau^3}.$$

*) Из большего числа таблиц интегралов Френеля можно рекомендовать, например, [24, 25, 73].

Снова интегрируя по частям и продолжая этот процесс, получим

$$\left.\begin{aligned}\mathscr{C}(w)&=\mathscr{C}(\infty)-\frac{1}{\pi w}\left[P(w)\cos\left(\frac{\pi}{2}w^2\right)-Q(w)\sin\left(\frac{\pi}{2}w^2\right)\right],\\ &\text{и аналогично}\\ \mathscr{S}(w)&=\mathscr{S}(\infty)-\frac{1}{\pi w}\left[P(w)\sin\left(\frac{\pi}{2}w^2\right)+Q(w)\cos\left(\frac{\pi}{2}w^2\right)\right],\end{aligned}\right\}\qquad(15)$$

где

$$\left.\begin{aligned}Q(w)&=1-\frac{1\cdot 3}{(\pi w^2)^2}+\frac{1\cdot 3\cdot 5\cdot 7}{(\pi w^2)^4}-\ldots,\\ P(w)&=\frac{1}{\pi w^2}-\frac{1\cdot 3\cdot 5}{(\pi w^2)^3}+\frac{1\cdot 3\cdot 5\cdot 7\cdot 9}{(\pi w^2)^5}-\ldots\end{aligned}\right\}\qquad(16)$$

Для вычисления интегралов $\mathscr{C}(\infty)$ и $\mathscr{S}(\infty)$ объединим их в один комплексный интеграл

$$\mathscr{C}(\infty)+i\mathscr{S}(\infty)=\int_0^\infty \exp\left(i\frac{\pi}{2}\tau^2\right)d\tau \qquad (17)$$

и введем новую переменную интегрирования

$$\zeta=\tau\sqrt{-\frac{i\pi}{2}}=\tau\frac{i-1}{2}\sqrt{\pi},\qquad \tau=-\zeta\frac{i+1}{\sqrt{\pi}}.$$

Вещественный путь интегрирования $0\leqslant\tau\leqslant\infty$ переходит в комплексной плоскости ζ в путь, лежащий вдоль линии, проходящей через начало координат под углом 45° к вещественной оси. Далее, легко заметить, что если интеграл берется вдоль линии, параллельной мнимой оси, то с увеличением расстояния от начала координат он начинает стремиться к нулю. В этом случае из теоремы о вычетах Коши следует, что интеграл, взятый вдоль любой наклонной линии, проходящей через начало координат, равен значению интеграла, взятого вдоль вещественной оси. Таким образом,

$$\mathscr{C}(\infty)+i\mathscr{S}(\infty)=\frac{i+1}{\sqrt{\pi}}\int_0^\infty \exp(-\zeta^2)\,d\zeta=\frac{i+1}{2}.$$

(Вещественный интеграл по ζ является хорошо известным интегралом ошибок Гаусса *), и его величина равна $\sqrt{\pi}/2$.) Следовательно,

$$\mathscr{C}(\infty)=\int_0^\infty \cos\left(\frac{\pi}{2}\tau^2\right)d\tau=\frac{1}{2},\qquad \mathscr{S}(\infty)=\int_0^\infty \sin\left(\frac{\pi}{2}\tau^2\right)d\tau=\frac{1}{2}. \qquad (18)$$

Соотношения (15) вместе с (16) и (18) выражают интегралы Френеля в виде рядов по отрицательным степеням w. Если w велико, эти расходящиеся (асимптотически) ряды обеспечивают хорошую аппроксимацию интегралов при учете небольшого числа членов разложения (см. приложение 3).

Поведение интегралов Френеля хорошо иллюстрируется изящным геометрическим построением Корню [75]. В качестве декартовых координат точки P берутся $\mathscr{C}$ и $\mathscr{S}$. Переменная w принимает все возможные значения, и поэтому точка P описывает некую кривую. Поскольку $\mathscr{S}(0)=\mathscr{C}(0)=0$, кривая проходит через начало координат, и поскольку

$$\mathscr{C}(-w)=-\mathscr{C}(w),\qquad \mathscr{S}(-w)=-\mathscr{S}(w), \qquad (19)$$

она антисимметрична относительно обеих осей. Если dl — элемент дуги нашей кривой, то

$$dl^2=d\mathscr{C}^2+d\mathscr{S}^2=\left[\left(\frac{d\mathscr{C}}{dw}\right)^2+\left(\frac{d\mathscr{S}}{dw}\right)^2\right](dw)^2=\left[\cos^2\left(\frac{\pi}{2}w^2\right)+\sin^2\left(\frac{\pi}{2}w^2\right)\right](dw)^2,$$

т. е.

$$(dl)^2=(dw)^2. \qquad (20)$$

*) См., например, [74].

Следовательно, если l измеряется в направлении увеличения w, то параметр w представляет длину дуги кривой, измеряемую от начала координат.

Пусть θ — угол между касательной к кривой и осью $\mathscr{C}$. Тогда

$$\operatorname{tg}\theta = \frac{d\mathscr{S}}{d\mathscr{C}} = \frac{\dfrac{d\mathscr{S}}{dw}}{\dfrac{d\mathscr{C}}{dw}} = \frac{\sin\left(\dfrac{\pi}{2}w^2\right)}{\cos\left(\dfrac{\pi}{2}w^2\right)} = \operatorname{tg}\left(\frac{\pi}{2}w^2\right),$$

т. е.

$$\theta = \frac{\pi}{2}w^2. \tag{21}$$

Следовательно, θ возрастает монотонно с увеличением $|w|$. Так как $\theta = 0$, когда $w = 0$, то в начале координат касательная к кривой совпадает с осью $\mathscr{C}$. Когда $w^2 = 1$, то $\theta = \pi/2$, и касательная перпендикулярна к оси $\mathscr{C}$. Если $w^2 = 2$, то $\theta = \pi$, и касательная к кривой снова параллельна оси $\mathscr{C}$, но ориентирована в отрицательном направлении. Согласно соотношениям (18) и (19) $\mathscr{C}(\infty) = -\mathscr{C}(-\infty) = 1/2$, $\mathscr{S}(\infty) = -\mathscr{S}(-\infty) = 1/2$, и обе ветви кривой приближаются к точкам P^+ и P^- с координатами (1/2, 1/2) и (—1/2,—1/2). Такая кривая называется *спиралью Корню* (рис. 8.35); она полезна при рассмотрении общих свойств дифракционных картин Френеля.

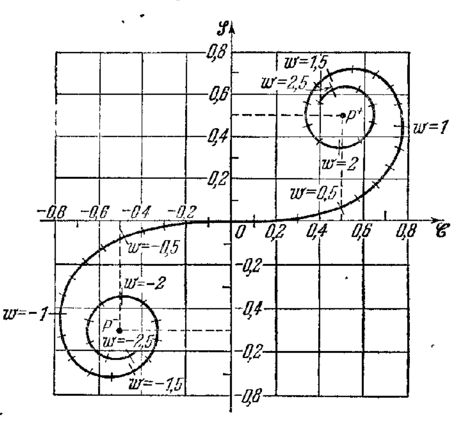


Рис. 8.35. Спираль Корню.

8.7.3. Дифракция Френеля на прямолинейном крае. Рассмотрим теперь дифракцию Френеля на полубесконечной плоскости, ограниченной острым прямолинейным краем. Это особенно важно для выяснения поведения поля вблизи границы геометрической тени. Ограничимся только случаем, когда линия P_0P, а также ее проекция (здесь ось x) на полуплоскость перпендикулярна краю (рис. 8.36). Если X — расстояние от края полуплоскости до начала координат (которое лежит на линии P_0P), то интегрирование производится по области

$$-\infty < \xi < X, \qquad -\infty < \eta < \infty,$$

или, переходя к переменным u и v —

$$-\infty < u < w, \qquad -\infty < v < \infty, \tag{22}$$

где

$$w = \sqrt{\frac{2}{\lambda}\left(\frac{1}{r'} + \frac{1}{s'}\right)} X\cos\delta. \tag{23}$$

Точка наблюдения P в зависимости от того, положительно ли X или отрицательно, лежит либо в освещенной области, либо в области геометрической тени.

Дифракционные интегралы (9) принимают тогда вид

$$\left.\begin{aligned} C &= b\int_{-\infty}^{w} du \int_{-\infty}^{+\infty} dv \left\{\cos\left(\frac{\pi}{2}u^2\right)\cos\left(\frac{\pi}{2}v^2\right) - \sin\left(\frac{\pi}{2}u^2\right)\sin\left(\frac{\pi}{2}v^2\right)\right\}, \\ S &= b\int_{-\infty}^{w} du \int_{-\infty}^{+\infty} dv \left\{\sin\left(\frac{\pi}{2}u^2\right)\cos\left(\frac{\pi}{2}v^2\right) + \cos\left(\frac{\pi}{2}u^2\right)\sin\left(\frac{\pi}{2}v^2\right)\right\}. \end{aligned}\right\} \tag{24}$$

Здесь нарушено условие, принятое при выводе (9), а именно условие, требующее, чтобы линейные размеры области интегрирования были малы по сравнению с расстояниями P_0O и OP. Поэтому для обоснования приближенной применимости этих формул и в данном случае необходимо более тщательное обсуждение остаточных членов. Однако сейчас мы не будем заниматься этим вопросом, так как в дальнейшем дифракация на полуплоскости будет разбираться более строгими методами (см. § 11.5).

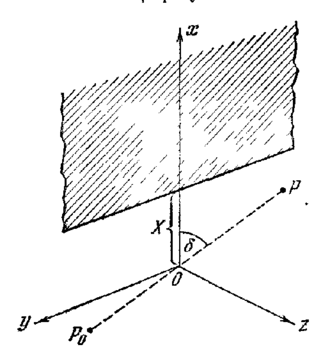


Рис. 8.36. К дифракции Френеля на прямолинейном крае.

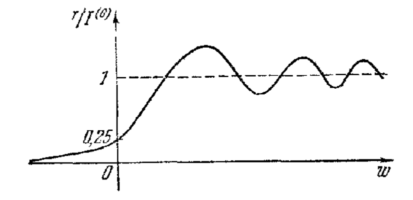


Рис. 8.37 Распределение интенсивности в картине дифракции Френеля на прямолинейном крае.

Из соотношений (18) и (19) имеем

$$\left.\begin{aligned}&\int_{-\infty}^{w}\cos\left(\frac{\pi}{2}\tau^2\right)d\tau=\int_{-\infty}^{0}+\int_{0}^{w}=\mathscr{C}(\infty)+\mathscr{C}(w)=\frac{1}{2}+\mathscr{C}(w),\\&\int_{-\infty}^{+\infty}\cos\left(\frac{\pi}{2}\tau^2\right)d\tau=1\end{aligned}\right\}\qquad(25)$$

и аналогично

$$\int_{-\infty}^{w}\sin\left(\frac{\pi}{2}\tau^2\right)d\tau=\frac{1}{2}+\mathscr{S}(w),\qquad\int_{-\infty}^{+\infty}\sin\left(\frac{\pi}{2}\tau^2\right)d\tau=1.\qquad(26)$$

Следовательно, (24) принимает вид

$$\left.\begin{aligned}C&=b\left\{\left[\frac{1}{2}+\mathscr{C}(w)\right]-\left[\frac{1}{2}+\mathscr{S}(w)\right]\right\},\\S&=b\left\{\left[\frac{1}{2}+\mathscr{C}(w)\right]+\left[\frac{1}{2}+\mathscr{S}(w)\right]\right\},\end{aligned}\right\}\qquad(27)$$

и подстановка в (4) дает окончательно выражение для интенсивности

$$I=\frac{1}{2}\left\{\left[\frac{1}{2}+\mathscr{C}(w)\right]^2+\left[\frac{1}{2}+\mathscr{S}(w)\right]^2\right\}I^{(0)},\qquad(28)$$

где

$$I^{(0)}=4\,|B|^2b^2=\frac{|A|^2}{(r'+s')^2}.\qquad(29)$$

Поведение функции интенсивности (28) можно установить, исследуя спираль Корню. Как мы видим, величина $2I/I^{(0)}$ равняется квадрату расстояния от точки w спирали Корню до «асимптотической» точки $P^-\left(-\frac{1}{2},\ -\frac{1}{2}\right)$. Таким образом, если точка наблюдения находится в освещенной области ($w>0$), то $I/I^{(0)}$ осциллирует с амплитудой, уменьшающейся по мере увеличения расстояния от края тени и, как это следует ожидать на основании геометрической оптики, асимптотически приближается к единице. Интенсивность максимальна не на границе геометрической тени, а на некотором расстоянии

от нее в освещенной области (рис. 8.37). На границе тени ($w=0$) имеем $I/I^{(0)}=1/4$. В области тени $I/I^{(0)}$ монотонно уменьшается до нуля. Эти выводы хорошо согласуются с экспериментальными результатами.

§ 8.8. Трехмерное распределение света вблизи фокуса

В п. 8.3.3 было показано, что распределение света в фокальной плоскости хорошо коррегированной линзы обусловлено по существу дифракцией Фраунгофера на ее оправе. В § 8.5 были подробно изучены картины дифракции Фраунгофера от отверстий различных форм. Для того чтобы получить более точное представление о структуре оптического изображения, следует изучить распределение света не только в геометрической фокальной плоскости, но и вблизи этой плоскости. Представление о трехмерном (Френель) распределении света вблизи фокуса имеет особенно важное значение для оценки величины допуска в требуемом положении плоскости изображения систем, формирующих изображение.

Особенности внефокальных монохроматических изображений точечного источника, полученных с помощью круглого отверстия, были впервые подробно рассмотрены в классическом труде Ломмеля *) [76]. Пользуясь интегралом Гюйгенса — Френеля ему удалось представить комплексное возмущение в виде сходящегося ряда функций Бесселя, а также экспериментально подтвердить явления, предсказанные на основании этих расчетов. Почти одновременно с Ломмелем Струве [78] опубликовал подобное, хотя и менее исчерпывающее исследование, посвященное дифракции на круглом отверстии. Он не получил таких подробных численных решений, но дал полезный метод приближенного расчета интенсивности вблизи границы геометрической тени, где ряды сходятся довольно медленно. Несколькими годами позже Шварцшильд [79] вывел асимптотическое приближение для точек наблюдения, находящихся на расстоянии многих длин волн от фокуса.

Исследования Ломмеля и Струве не привлекли особого внимания, и в 1909 г. эта задача снова была исследована Дебаем [80]. Дебай установил определенные общие особенности дифракционного поля и вблизи и вдали от фокуса. Сравнительно недавно исследования перечисленных выше авторов были расширены и были опубликованы диаграммы, показывающие подробную структуру поля в этой комплексной области. Результаты проведенных исследований получили широкое экспериментальное подтверждение как в оптическом, так и в микроволновом (короткие радиоволны) диапазонах.

Рассматривая распределение света вблизи фокуса, мы будем основываться на исследованиях Ломмеля и Струве, но для удобства и наглядности начнем с интегрального представления поля в виде, предложенном Дебаем.

8.8.1. Вычисление дифракционного интеграла в функциях Ломмеля. Рассмотрим сферическую монохроматическую волну, выходящую из круглого отверстия и сходящуюся в осевой фокальной точке O. Рассмотрим возмущение $U(P)$ в произвольной точке P близ O. Положение точки P относительно O определяется вектором $\mathbf{R}$. Предполагается, что расстояние $R=OP$ и радиус $a\,(\gg\lambda)$ отверстия малы по сравнению с радиусом $f=CO$ волнового фронта W, который в какой-то момент заполняет это отверстие (рис. 8.38).

Обозначим через s расстояние от точки наблюдения P до точки Q на W, а через A/f — амплитуду в точке Q падающей волны; тогда, применяя принцип Гюйгенса — Френеля, получим

$$U(P)=-\frac{i}{\lambda}\,\frac{A\exp(-ikf)}{f}\iint\frac{\exp(iks)}{s}\,dS. \qquad (1)$$

*) Более поздняя статья [77] касается дифракции на щели, на непрозрачной полоске и прямолинейном крае.

Поскольку рассматриваются лишь небольшие углы падения, вариации коэффициента наклона по волновому фронту пренебрежимо малы. Обозначая через $\mathbf{q}$ единичный вектор в направлении OQ, находим с хорошим приближением

$$s-f=-\mathbf{q}\cdot\mathbf{R}. \tag{2}$$

Далее элемент dS волнового фронта равняется

$$dS=f^2\,d\Omega, \tag{3}$$

где $d\Omega$ — элемент телесного угла, под которым dS виден из точки O. Без заметной ошибки можно заменить в знаменателе подынтегрального выражения s на f, и тогда уравнение (1) запишется в виде

$$U(P)=-\frac{i}{\lambda}A\iint\limits_{\Omega}\exp(-ik\mathbf{q}\cdot\mathbf{R})\,d\Omega. \tag{4}$$

Теперь интегрирование производится по всему телесному углу Ω, под которым отверстие видно из фокуса. Уравнение (4) представляет собой *интеграл Дебая* и выражает поле как результат суперпозиции плоских волн, распространяющихся во всех направлениях (определяемых векторами $\mathbf{q}$), попадающих в Ω.

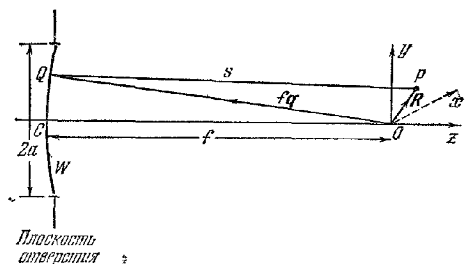


Рис. 8.38. К дифракции на круглом отверстии сходящейся сферической волны.
$u=\frac{2\pi}{\lambda}\left(\frac{a}{f}\right)^2 z$, $v=\frac{2\pi}{\lambda}\left(\frac{a}{f}\right)\sqrt{x^2+y^2}$

Перед тем как вычислять интеграл Дебая, отметим интересное обстоятельство, будучи суммой элементарных решений (плоских волн), он представляет строгое решение волнового уравнения и в предельном случае $f\to\infty$ (отверстие на бесконечном расстоянии) справедливо во всем пространстве. Конечно, (4) нельзя считать строгим решением нашей исходной задачи, так как здесь не учтена природа экрана, а точные граничные условия аппроксимируются граничными условиями дифракционной теории Кирхгофа. Точное решение нашей задачи должно содержать не только вклады от плоских волн, распространяющихся в направлении падающих геометрических лучей, но и вклады от волн, распространяющихся во всех возможных направлениях*). Однако при выполнении упомянутых выше условий ($f\gg a\gg\lambda$, $f\gg R$) значительны только вклады от волн, учтенных в уравнении (4).

Для вычисления (4) представим подынтегральное выражение в более явном виде. Примем за начало декартовых координат точку O и за ось z — линию Oz. Пусть (x, y, z) — координаты точки P, а (ξ, η, ζ) — координаты точки Q. Положим, кроме того, что

$$\xi=a\rho\sin\theta,\qquad \eta=a\rho\cos\theta,\qquad x=r\sin\psi,\qquad y=r\cos\psi. \tag{5}$$

Так как точка Q лежит на сферическом волновом фронте W, имеем

$$\zeta=-\sqrt{f^2-a^2\rho^2}=-f\left[1-\frac{1}{2}\frac{a^2\rho^2}{f^2}+\ldots\right] \tag{6}$$

Следовательно,

$$\mathbf{q}\cdot\mathbf{R}=\frac{x\xi+y\eta+z\zeta}{f}=\frac{a\rho r\cos(\theta-\psi)}{f}-z\left[1-\frac{1}{2}\frac{a^2\rho^2}{f^2}+\ldots\right]. \tag{7}$$

*) Это соответствует представлению поля через так называемый угловой спектр плоских волн (см. п. 11.4.2).

Полезно на этом этапе ввести безразмерные переменные u и v, которые вместе с ψ определяют положение P, а именно

$$u=\frac{2\pi}{\lambda}\left(\frac{a}{f}\right)^2 z, \qquad v=\frac{2\pi}{\lambda}\left(\frac{a}{f}\right) r=\frac{2\pi}{\lambda}\frac{a}{f}\sqrt{x^2+y^2}. \tag{8}$$

Заметим, что точка P лежит в прямом пучке света или в геометрической тени, смотря по тому, будет ли $|v/u|\lessgtr 1$.

Из (7) и (8) следует, что если члены с ρ/f в степени, превышающей вторую, пренебрежимо малы по сравнению с единицей, то

$$k\mathbf{q}\cdot\mathbf{R}=v\rho\cos(\theta-\psi)-\left(\frac{f}{a}\right)^2 u+\frac{1}{2}u\rho^2. \tag{9}$$

Кроме того, элемент телесного угла равен

$$d\Omega=\frac{dS}{f^2}=\frac{a^2\rho\,d\rho\,d\theta}{f^2}. \tag{10}$$

Поэтому (4) принимает вид

$$U(P)= -\frac{i}{\lambda}\frac{a^2A}{f^2}\exp\left[i\left(\frac{f}{a}\right)^2 u\right]\int_0^1\int_0^{2\pi}\exp\left\{-i\left[v\rho\cos(\theta-\psi)+\frac{1}{2}u\rho^2\right]\right\}\rho\,d\rho\,d\theta. \tag{11}$$

Интеграл по θ мы уже встречали, рассматривая дифракцию Фраунгофера на круглом отверстии (см. п. 8.5.2). Он равен $2\pi J_0(v\rho)$, где $J_0(v\rho)$ — функция Бесселя нулевого порядка. Следовательно, последнее соотношение можно записать как

$$U(P)=-\frac{2\pi i a^2 A}{\lambda f^2}\exp\left\{i\left(\frac{f}{a}\right)^2 u\right\}\int_0^1 J_0(v\rho)\exp\left(-\frac{1}{2}iu\rho^2\right)\rho\,d\rho. \tag{12}$$

Рассмотрим по отдельности вещественную и мнимую части интеграла. Положим, что

$$2\int_0^1 J_0(v\rho)\exp\left(-\frac{1}{2}iu\rho^2\right)\rho\,d\rho=C(u,v)-iS(u,v), \tag{13}$$

где

$$\left.\begin{aligned} C(u,v)&=2\int_0^1 J_0(v\rho)\cos\left(\frac{1}{2}u\rho^2\right)\rho\,d\rho,\\ S(u,v)&=2\int_0^1 J_0(v\rho)\sin\left(\frac{1}{2}u\rho^2\right)\rho\,d\rho. \end{aligned}\right\} \tag{14}$$

Эти интегралы можно вычислить в *функциях Ломмеля*

$$\left.\begin{aligned} U_n(u,v)&=\sum_{s=0}^{\infty}(-1)^s\left(\frac{u}{v}\right)^{n+2s}J_{n+2s}(v),\\ V_n(u,v)&=\sum_{s=0}^{\infty}(-1)^s\left(\frac{v}{u}\right)^{n+2s}J_{n+2s}(v), \end{aligned}\right\} \tag{15}$$

введенных им для этой цели *). Используя соотношение (8.5.11)

$$\frac{\partial}{\partial x}\left[x^{n+1}J_{n+1}(x)\right]=x^{n+1}J_n(x),$$

*) Полное обсуждение этих функций приводится в [76, 77], а также в книгах [25, 81, 82].

можно записать $C(u, v)$ после интегрирования по частям в виде

$$C(u, v)=\frac{2}{v}\int_0^1 \frac{d}{d\rho}[\rho J_1(v\rho)]\cos(^1/_2 u\rho^2)\,d\rho=$$
$$=\frac{2}{v}\left[J_1(v)\cos{}^1/_2 u+u\int_0^1 \rho^2 J_1(v\rho)\sin(^1/_2 u\rho^2)\,d\rho\right]. \qquad (16)$$

Используя опять соотношение (8.5.11), интегрируя по частям и продолжая этот процесс, получим

$$C(u, v)=\frac{\cos(^1/_2 u)}{^1/_2 u}\left[\frac{u}{v}J_1(v)-\left(\frac{u}{v}\right)^3 J_3(v)+\ldots\right]+$$
$$+\frac{\sin(^1/_2 u)}{^1/_2 u}\left[\left(\frac{u}{v}\right)^2 J_2(v)-\left(\frac{u}{v}\right)^4 J_4(v)+\ldots\right]=$$
$$=\frac{\cos(^1/_2 u)}{^1/_2 u}U_1(u, v)+\frac{\sin(^1/_2 u)}{^1/_2 u}U_2(u, v). \qquad (17а)$$

Таким же способом найдем

$$S(u, v)=\frac{\sin(^1/_2 u)}{^1/_2 u}U_1(u, v)-\frac{\cos(^1/_2 u)}{^1/_2 u}U_2(u, v). \qquad (17б)$$

Эти формулы справедливы во всех точках близ фокуса, но удобны для вычислений, только если $|u/v|<1$, т. е. когда точка наблюдения лежит в геометрической тени. При $|u/v|>1$, т. е. когда точка наблюдения находится в освещенной области, более целесообразно применять разложения, содержащие положительные степени v/u. Их можно вывести аналогичным способом с помощью интегрирования по частям относительно тригонометрического члена. Первый этап дает

$$C(u, v)=\frac{2}{u}\int_0^1 J_0(v\rho)\frac{d}{d\rho}[\sin(^1/_2 u\rho^2)]\,d\rho=$$
$$=\frac{2}{u}\left[J_0(v)\sin{}^1/_2 u-v\int_0^1 J_1(v\rho)\sin(^1/_2 u\rho^2)\rho\,d\rho\right]; \qquad (18)$$

здесь использовано соотношение (8.5.15) $\frac{\partial}{\partial x}[x^{-n}J_n(x)]=-x^{-n}J_{n+1}(x)$. Снова интегрируя по частям и используя последнее соотношение совместно с хорошо известной формулой (которая выводится из разложения в ряд $J_n(x)$)

$$\lim_{x\to 0}\frac{J_n(x)}{x^n}=\frac{1}{2^n n!}, \qquad (19)$$

получим

$$C(u, v)=\frac{\sin(^1/_2 u)}{^1/_2 u}\left[J_0(v)-\left(\frac{v}{u}\right)^2 J_2(v)+\ldots\right]-$$
$$-\frac{\cos(^1/_2 u)}{^1/_2 u}\left[\left(\frac{v}{u}\right)J_1(v)-\left(\frac{v}{u}\right)^3 J_3(v)+\ldots\right]+$$
$$+\frac{2}{u}\left[\frac{v^2}{2u}-\frac{1}{3!}\left(\frac{v^2}{2u}\right)^3+\ldots\right].$$

Ряд, написанный в первых двух строчках, содержит две функции Ломмеля V_n, а третья строчка представляет собой разложение в ряд $\sin(v^2/2u)$. Следовательно,

$$C(u, v)=\frac{2}{u}\sin\frac{v^2}{2u}+\frac{\sin(^1/_2 u)}{^1/_2 u}V_0(u, v)-\frac{\cos(^1/_2 u)}{^1/_2 u}V_1(u, v). \qquad (20а)$$

Аналогичным образом получим для другого интеграла выражение

$$S(u, v)=\frac{2}{u}\cos\frac{v^2}{2u}-\frac{\cos(^1/_2 u)}{^1/_2 u}V_0(u, v)-\frac{\sin(^1/_2 u)}{^1/_2 u}V_1(u, v). \qquad (20б)$$

Этим завершается формальное решение нашей задачи. Далее мы обсудим некоторые выводы из него.

8.8.2. **Распределение интенсивности.** Согласно (12), (13), (17) и (20) интенсивность $I=|U|^2$ близ фокуса определяется двумя эквивалентными выражениями, а именно

$$I(u, v)=\left(\frac{2}{u}\right)^2[U_1^2(u, v)+U_2^2(u, v)]\,I_0 \tag{21а}$$

и

$$I(u, v)=\left(\frac{2}{u}\right)^2\left[1+V_0^2(u, v)+V_1^2(u, v)-2V_0(u, v)\cos\left\{{}^1/_2\left(u+\frac{v^2}{u}\right)\right\}-\right.$$
$$\left.-2V_1(u, v)\sin\left\{{}^1/_2\left(u+\frac{v^2}{u}\right)\right\}\right]I_0, \tag{21б}$$

где

$$I_0=\left(\frac{\pi a^2 A}{\lambda f^2}\right)^2 \tag{22}$$

— интенсивность в геометрическом фокусе $u=v=0$.

Из (15) следует, что $U_1(-u, v)=-U_1(u, v)$, $U_2(-u, v)=U_2(u, v)$, $V_0(-u, v)=V_0(u, v)$, $V_1(-u, v)=-V_1(u, v)$.

Таким образом, $I(u, v)$ остается неизменной при замене u на $-u$. Значит, *близ фокуса распределение интенсивности симметрично относительно геометрической фокальной плоскости.* Конечно, распределение интенсивности симметрично также и относительно оси $v=0$.

По формулам (21) Ломмель рассчитал распределение интенсивности для ряда выбранных плоскостей, находящихся вблизи фокуса; он экспериментально подтвердил также некоторые свои расчеты *). Линии равной интенсивности (называемые *изофотами*) вблизи фокуса, построенные по данным Ломмеля, приведены **) на рис. 8.39.

Особый интерес представляет «трубчатая» структура светлой центральной части дифракционного изображения, хорошо заметная на рисунке; ее существование постулировалось на основании экспериментальных данных еще в 1894 г. Тейлором [92]. Именно этой структурой изображения определяется допуск в положении плоскости изображения в системах, формирующих изображение.

Ниже рассматриваются несколько интересных частных случаев.

а. Интенсивность в геометрической фокальной плоскости. Для точек, лежащих в геометрической фокальной плоскости, $u=0$, и уравнение (21а) приводится к

$$I(0, v)=4\lim_{u\to 0}\left[\frac{U_1^2(u, v)+U_2^2(u, v)}{u^2}\right]I_0. \tag{23}$$

Из уравнений, определяющих функции U_n, следует, что

$$\lim_{u\to 0}\left[\frac{U_1(u, v)}{u}\right]=\frac{J_1(v)}{v}, \quad \lim_{u\to 0}\left[\frac{U_2(u, v)}{u}\right]=0, \tag{24}$$

поэтому

$$I(0, v)=\left[\frac{2J_1(v)}{v}\right]^2 I_0. \tag{25}$$

*) Сходные эксперименты описаны также в [83, 84]. Аналогичные опыты с микроволнами изложены в [85] для отверстия круглой формы и в [86] — для прямоугольного отверстия

**) Похожие, но менее детальные графики опубликованы в [87] Они воспроизведены в некоторых книгах с ошибками (неправильное положение геометрической тени и перепутанные оси) Другой тип графиков, по существу соответствующий рис 8 39, приведен Цернике и Нижбером в [89] Они исходили из другого, но эквивалентного разложения (см (9 4 12)) дифракционного интеграла Графики, рассчитанные на основе электромагнитной теории и показывающие распределение электрической энергии и потока энергии, приведены в [89а].

Аналогичные диаграммы для кольцевого отверстия опубликованы Линфутом и Вольфом [90]. Некоторые применения исследований Ломмеля к дифракции на системе концентрических кольцеобразных отверстий приведены в [91].

Как и следовало ожидать, мы получили формулу Эйри (8.5.14) для дифракции Фраунгофера на круглом отверстии.

б. Распределение интенсивности вдоль оси. Для точек, лежащих на оси, $v=0$, и обе функции V_n, входящие в выражение (21б), запишутся в виде

$$V_0(u,\ 0)=1,\qquad V_1(u,\ 0)=0.$$

Следовательно,

$$I(0,\ v)=\frac{4}{u^2}\left[2-2\cos\left({}^1/_2\, u\right)\right]I_0=\left(\frac{\sin(u/4)}{u/4}\right)^2 I_0. \tag{26}$$

Таким образом, интенсивность вдоль оси характеризуется функцией $(\sin x/x)^2$,

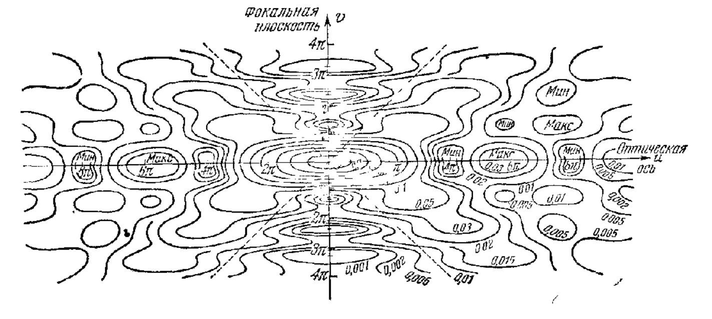


Рис. 8.39. Изофоты (линии интенсивности $I(u,\ v)$) в меридиональной плоскости вблизи фокуса сходящейся сферической волны, дифрагировавшей на круглом отверстии (из [87]).

Интенсивность в фокусе нормирована к единице. Пунктирные линии показывают границу геометрической тени. При вращении рисунка вокруг оси u минимумы на оси v вырисовывают темные кольца Эйри.

рассмотренной нами в п. 8.5.1 в связи с дифракцией Фраунгофера на прямоугольном отверстии. Первый нуль интенсивности на оси получается при $u/4\equiv 2\pi a^2 z/\lambda f^2=\pm\pi$, т. е. на расстоянии $z=\pm f^2\lambda/2a^2$ от фокуса.

Обычно потеря интенсивности в центральном пятне изображения, достигающая 20%, считается допустимой. Так как $\left(\frac{\sin(u/4)}{u/4}\right)^2$ уменьшается на эту величину при смещении плоскости изображения от положения $u=0$ до положения $u\approx 3{,}2$, то, следовательно, допуск на положение фокальной плоскости Δz равен приблизительно

$$\Delta z=\pm 3{,}2\frac{\lambda}{2\pi}\left(\frac{f}{a}\right)^2\approx\pm{}^1/_2\left(\frac{f}{a}\right)^2\lambda. \tag{27}$$

Например, при использовании пучка $f/10$ $(f/a=20)$ для света с длиной волны $\lambda=5\cdot 10^{-5}$ *см* этот допуск составляет около $\pm 0{,}5\cdot 20^2\cdot 5\cdot 10^{-5}$ *см* $=\pm 0{,}1$ *мм*.

в. Интенсивность вдоль границы геометрической тени. Для точек на границе геометрической тени $u=\pm v$. Распределение интенсивности симметрично относительно геометрической фокальной плоскости, и поэтому, не теряя общности, можно считать u положительным. Тогда функция U_n сведется к

$$U_1(u,\ u)=\sum_{s=0}^{\infty}(-1)^s J_{2s+1}(u),\qquad U_2(u,\ u)=\sum_{s=0}^{\infty}(-1)^s J_{2s+2}(u). \tag{28}$$

Напомним хорошо известное тождество Якоби (см., например, [25])

$$\left.\begin{aligned}\cos(u\cos\theta)&=J_0(u)+2\sum_{s=1}^{\infty}(-1)^s J_{2s}(u)\cos 2s\theta,\\ \sin(u\cos\theta)&=2\sum_{s=0}^{\infty}(-1)^s J_{2s+1}(u)\cos(2s+1)\theta.\end{aligned}\right\} \tag{29}$$

Полагая $\theta = 0$, из сравнения с (28) находим

$$U_1(u, u) = {}^1/_2 \sin u, \qquad U_2(u, u) = {}^1/_2 [J_0(u) - \cos u], \tag{30}$$

и (21а) сводится к

$$I(u, u) = \frac{1 - 2J_0(u)\cos u + J_0^2(u)}{u^2} I_0. \tag{31}$$

График этой функции показан на рис. 8.40.

8.8.3. Суммарная интенсивность. Представляет интерес определить часть L от усредненной по времени полной энергии, проходящей через небольшое круглое отверстие заданного радиуса r_0 около осевой точки плоскости $u = \text{const}$ изображения. Если

$$E = \pi a^2 \left(\frac{A}{f}\right)^2 \tag{32}$$

— полная энергия, падающая на отверстие в единицу времени, то искомая часть энергии равняется

$$L(u, v_0) = \frac{1}{E} \int_0^{v_0} \int_0^{2\pi} I(u, v)\, r\, dr\, d\psi =$$

$$= \frac{1}{2I_0} \int_0^{v_0} I(u, v)\, v\, dv, \tag{33}$$

где

$$v_0 = \frac{2\pi}{\lambda} \left(\frac{a}{f}\right) r_0. \tag{34}$$

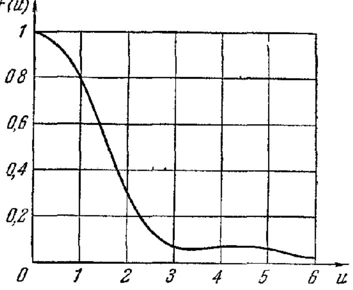


Рис. 8.40. Изменение интенсивности вдоль границы геометрической тени. График функции

$$F(u) = \frac{1 - 2J_0(u)\cos u + J_0^2(u)}{u^2}$$

Подставив выражения (21) Ломмеля для интенсивности в (33), можно разложить интеграл в ряд, содержащий функции Бесселя. Так как эта процедура слишком длинна, мы приведем только окончательный результат, полученный Вольфом [93]. Асимптотические приближения для $L(u, v)$ получены Фокке [94].

Здесь мы снова получим два формально различных выражения, одно, удобное для вычислений, когда граница маленького круглого отверстия находится в геометрической тени, другое — удобное при расчете, когда это отверстие находится в прямом пучке света. Отбрасывая индекс нуль, т. е написав v вместо v_0, получим в первом случае ($|v/u| \geqslant 1$)

$$L(u, v) = 1 - \sum_{s=0}^{\infty} \frac{(-1)^s}{2s+1} \left(\frac{u}{v}\right)^{2s} Q_{2s}(v), \tag{35а}$$

где

$$Q_{2s}(v) = \sum_{p=0}^{2s} (-1)^p [J_p(v) J_{2s-p}(v) + J_{p+1}(v) J_{2s+1-p}(v)]. \tag{36}$$

Во втором случае ($|v/u| \leqslant 1$) находим

$$L(u, v) = \left(\frac{v}{u}\right)^2 \left[1 - \sum_{s=0}^{\infty} \frac{(-1)^s}{2s+1} \left(\frac{v}{u}\right)^{2s} Q_{2s}(v)\right] -$$

$$- \frac{4}{u} \left[Y_1(u, v) \cos \frac{1}{2}\left(u + \frac{v^2}{u}\right) + Y_2(u, v) \sin \frac{1}{2}\left(u + \frac{v^2}{u}\right)\right], \tag{35б}$$

где величины Q задаются (36), а Y_1 и Y_2 определяются выражениями

$$Y_n(u, v) = \sum_{s=0}^{\infty} (-1)^s (n+2s) \left(\frac{v}{u}\right)^{n+2s} J_{n+2s}(v) = \frac{1}{2}\left[\frac{v^2}{u} V_{n-1}(u, v) + u V_{n+1}(u, v)\right]. \tag{37}$$

На рис. 8 41 показано семейство кривых $L(u, v)$, рассчитанных по этим формулам. В известном смысле эти кривые можно считать аналогами лучей в геометрической оптике

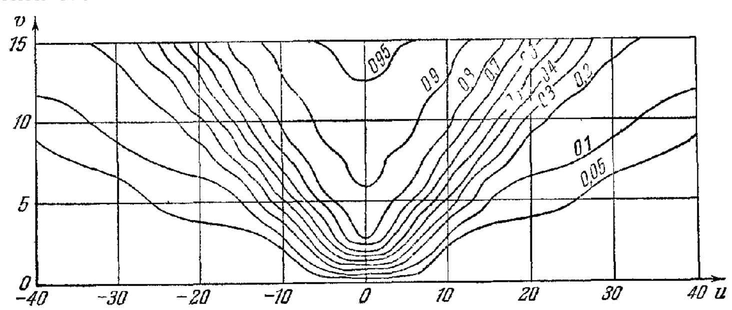


Рис 8 41 Семейство кривых $L(u, v)$, показывающих часть полной энергии, попадающей внутрь небольшого кружка с центром на оси в выбранной плоскости изображения u =const [93]

Следует отметить, что в частном случае, когда плоскость изображения совпадает с фокальной плоскостью ($u=0$), соотношение (35а) принимает вид

$$L(0, v)=1-Q_0(v)=1-J_0^2(v)-J_1^2(v), \quad (38)$$

что находится в согласии с формулой Рэлея (8 5 18)

Особый интерес представляет случай, когда кружок, по которому берется интеграл (33), совпадает с сечением геометрического конуса лучей Тогда $|v/u|=1$, ряды в (35а) и (35б) можно просуммировать, и мы получим [93]

$$L(u, u)=1-J_0(u)\cos u-J_1(u)\sin u. \quad (39)$$

Следовательно, выражение

$$\varepsilon(u)=J_0(u)\cos u+J_1(u)\sin u \quad (40)$$

дает часть полной энергии, попадающую в геометрическую тень, в плоскости $u=$ = const График функции $\varepsilon(u)$ приведен на рис 8 42, она уменьшается не монотонно, а имеет максимумы (кроме $u=0$) при $J_1(u)=0$ и минимумы при $\sin u=0\,(u\neq 0)$.

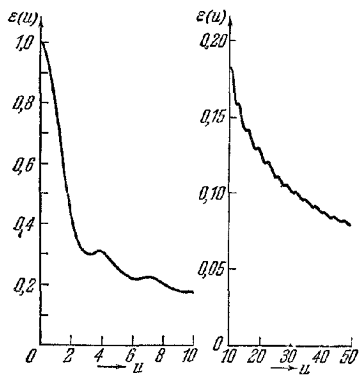


Рис 8 42 Часть $\varepsilon(u)$ от полной энергии, приходящейся на область, занимаемую геометрической тенью [93]

8.8.4. Фазовые соотношения. Рассмотрим, наконец, поведение фазы возмущения в области близ фокуса Согласно (12) и (13) фаза (не считая аддитивного члена ($-\omega t$)) определяется выражением *)

$$\varphi(u, v)=(f/a)^2 u-\chi(u, v)-\pi/2 \quad (\mathrm{mod}\, 2\pi), \quad (41)$$

где

$$\cos\chi=\frac{C}{\sqrt{C^2+S^2}}, \quad \sin\chi=\frac{S}{\sqrt{C^2+S^2}}. \quad (42)$$

В (42) квадратный корень берется со знаком плюс.

*) Символ mod 2π справа от уравнения означает, что обе части уравнения определены с точностью до аддитивной постоянной $2m\pi$, где m — целое число.

Заметим, что, в отличие от распределения интенсивности, распределение фаз нельзя выразить только через u и v, так как оно имеет структуру, зависящую от угловой апертуры геометрического пучка лучей. Далее каждая «ветвь» многозначной функции $\varphi(u, v)$ непрерывна по u и v во всех точках, где интенсивность не исчезающе мала, а в точках с нулевой интенсивностью она неопределенна. В фокусе, где $u = v = 0$, одно из ее значений равно $-\pi/2$.

Синфазные поверхности (поверхность, где $\varphi = \mathrm{const}$) — это, конечно, поверхности вращения относительно оси u. Покажем, что такие поверхности обладают еще и другой симметрией, определяемой соотношением

$$\varphi(-u, v) + \varphi(u, v) = -\pi \qquad (\mathrm{mod}\, 2\pi). \tag{43}$$

Из (14) имеем:

$$C(-u, v) = C(u, v), \qquad S(-u, v) = -S(u, v). \tag{44}$$

Тогда из (43) следует, что

$$\cos\chi(-u, v) = \cos\chi(u, v), \qquad \sin\chi(-u, v) = -\sin\chi(u, v), \tag{45}$$

а из уравнения (42) с учетом (43) и (45) находим

$$\begin{aligned}\cos[\varphi(-u, v) + \varphi(u, v)] &= \cos[-\chi(-u, v) + \chi(u, v) - \pi] = \\ &= -\cos[\chi(-u, v) + \chi(u, v)] = \\ &= -\cos\chi(-u, v)\cos\chi(u, v) + \sin\chi(-u, v)\sin\chi(u, v) = \\ &= -\cos^2\chi(u, v) - \sin^2\chi(u, v) = -1.\end{aligned} \tag{46}$$

Следовательно, «зеркальное изображение» плоскостью $u = 0$ любой синфазной поверхности $\varphi = \varphi_0$ также является синфазной поверхностью $\varphi = -\pi - \varphi_0$.

На рис. 8.43 показаны профили синфазных поверхностей гомоцентрического пучка лучей с $f/2$. Вдали от фокуса синфазные поверхности совпадают

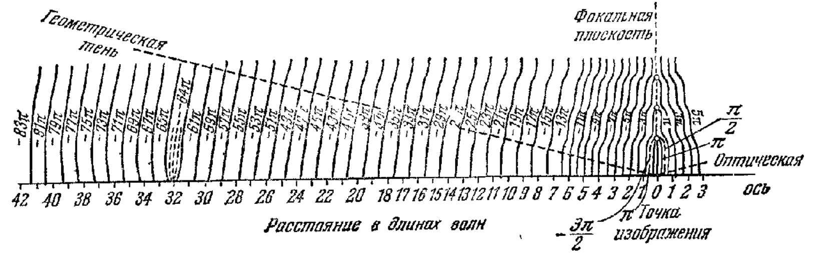


Рис. 8.43. Профили синфазных поверхностей $\varphi(u, v) = \mathrm{const}$ вблизи геометрического фокуса [95]

$\lambda = 5\cdot 10^{-5}$ *см*, $a = 2{,}5$ *см*, $f = 10$ *см*. Константы для поверхностей отличаются на π от величин, приведенных в статье Фарнелла. Это сделано для того, чтобы согласовать рисунок с нашим текстом, где, согласно (51), фаза в фокусе принята за $-\pi/2$, тогда как у Фарнелла она принята за $+\pi/2$.

со сферическими волновыми фронтами геометрической оптики, но все сильнее и сильнее деформируются по мере приближения к фокальной области. В непосредственной близости от геометрической фокальной плоскости они показаны на рис. 8.44. Как мы видим, вблизи фокуса синфазные поверхности достаточно плоски, но расположены в $1 - a^2/4f^2$ раз дальше, чем это было бы в параллельном пучке света с такой же длиной волны. Более того, на каждой синфазной поверхности интенсивность непостоянна (см. рис. 8.39). В непосредственной близости к темным кольцам Эйри (обозначенным на рис. 8.44 через R_1 и R_2) поведение синфазных поверхностей более сложно. При перемещении точки наблюдения в направлении v от геометрического фокуса фаза остается постоянной в промежутке между любыми двумя соседними темными кольцами, но при пересечении каждого кольца внезапно изменяется на π. Справедливость последнего

утверждения вытекает из выражения (24) для комплексного возмущения в геометрической фокальной плоскости, т. е.

$$U(0,\ v) = \frac{1}{i}\,\frac{2J_1(v)}{v}\sqrt{I_0}. \tag{47}$$

Так как последнее выражение чисто мнимое при всех значениях v, то фаза $U(0,\ v)$ может равняться только $-\pi/2$ или $+\pi/2$ (mod 2π), и поскольку ее знак изменяется при пересечении каждого темного кольца, то фаза изменяется на π скачком. Такой же скачок наблюдается и при переходе через каждую осевую точку с нулевой интенсивностью.

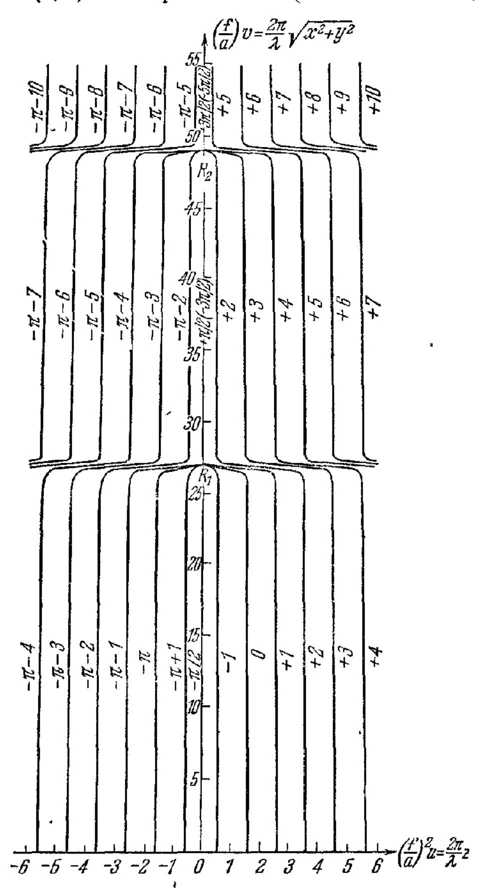


Рис. 8.44. Профили синфазных поверхностей в непосредственной близости от геометрической фокальной плоскости гомоцентрического пучка с $f/3{,}5$ [87]. OR_1 и OR_2 — радиусы первого и второго темных колец Эйри

Интересно также рассмотреть изменение фазы поля при движении точки наблюдения вдоль каждого луча, проходящего через фокус. Это удобно сделать, сравнивая такие изменения с изменениями фазы сферической волны, сходящейся в фокус в полупространстве $z<0$ и расходящейся затем в полупространстве $z>0$. Приняв фазу $\tilde{\varphi}$ этой волны сравнения в фокусе за нуль, получим

$$\left.\begin{aligned} \tilde{\varphi}(u,\ v) &= -kR, \quad \text{если} \quad u \leqslant 0, \\ \tilde{\varphi}(u,\ v) &= +kR, \quad \text{если} \quad u \geqslant 0, \end{aligned}\right\} \tag{48}$$

где, как и раньше, $R = \sqrt{x^2+y^2+z^2} > 0$ — расстояние между точкой наблюдения и фокусом. Разность

$$\delta(u,\ v) = \varphi(u,\ v) - \tilde{\varphi}(u,\ v) \tag{49}$$

называется *аномалией фазы*.

Из (43), (48) и (49) следует, что

$$\delta(-u,\ v) + \delta(u,\ v) = -\pi \quad (\mathrm{mod}\ 2\pi), \tag{50}$$

тогда как в самом фокусе

$$\delta(0,\ 0) = \varphi(0,\ 0) = -\frac{\pi}{2} \quad (\mathrm{mod}\ 2\pi). \tag{51}$$

Аномалия фазы вдоль выбранных лучей гомоцентрического пучка с $f/3{,}5$, проходящих через фокус, показана на рис. 8.45

Как видно на рисунке, при прохождении через фокус вдоль любого луча, кроме аксиального, δ испытывает быстрое, но непрерывное изменение на π. Этому эффекту, наблюдавшемуся много лет тому назад Гуи [96], были посвя-

щены многочисленные исследования (обзор литературы приведен в [97], ссылки на более поздние исследования можно найти в [87]). Вдоль оси, однако, аномалия фазы ведет себя особым образом, она изменяется периодически от 0 до $-\pi$.

Рассматривая асимптотическое приближение интеграла Гюйгенса — Френеля при больших k (коротких длинах волн), нетрудно показать, что фаза скачком меняется на $\pi/2$, когда свет проходит через каждый из двух главных центров кривизны соответствующих волновых фронтов [12, 82]. Разобранный выше

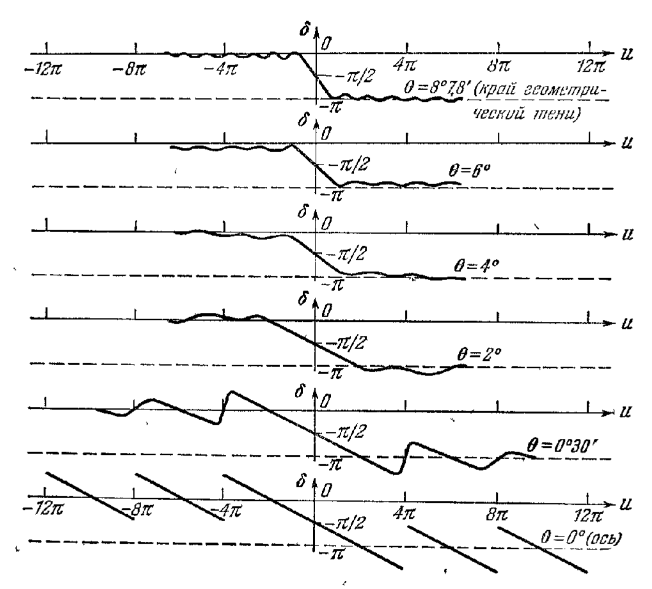


Рис 8.45. Аномалия фазы δ вдоль геометрических лучей, проходящих через фокус гомоцентрического пучка с $f/3,5$ [87].
θ — угол наклона луча к оси.

случай соответствует особому положению, когда оба центра кривизны совпадают. Поэтому изменение фазы на половину периода соответствует решению в рамках геометрической оптики (обсуждение этого вопроса см. в [98]). На рис. 8.45 показано, как разрывы в геометрической оптике превращаются в непрерывные переходы, если учитывается конечность длин волн. Наконец, сошлемся на статью Фарнела [99], в которой описано экспериментальное исследование структуры распределения фазы в области фокуса линзы. Это исследование, проведенное в микроволновом диапазоне, дало очень хорошее совпадение с предсказаниями теории.

§ 8.9. Граничная дифрагировавшая волна

Если рассматривать край дифракционного отверстия (или препятствия из точек, лежащих внутри геометрической тени), то он кажется светящимся. Это было известно уже Юнгу [100], пытавшемуся еще до Френеля теоретически объяснить дифракцию, исходя из волновой теории. Юнг полагал, что падающий свет испытывает род отражения на краю дифрагирующего тела и рассматривал дифракционную картину как результат интерференции падающей волны и

отраженной «граничной волны». Однако представления Юнга были выражены только качественно и не получили широкого признания.

Наличие элементов истины в теории Юнга стало очевидным только после того, как Зоммерфельд получил в 1894 г. строгое решение задачи о дифракции плоских волн на плоском полубесконечном отражающем экране (см. § 11.5). Это решение показывает, что в геометрической тени свет распространяется в виде цилиндрической волны, которая кажется исходящей от края экрана, тогда как в освещенной области она представляется суперпозицией цилиндрической и исходной падающей волн.

Возникает вопрос, можно ли и в более общем случае объяснить дифракцию как совместное действие падающей и граничной волн. Этот вопрос был исследован Магги [101] до появления работы Зоммерфельда. Но его результаты,

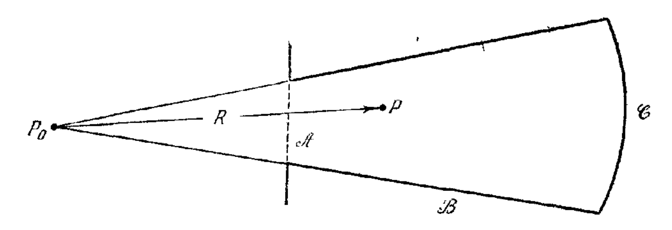


Рис 8.46. К выводу выражении для граничной дифрагировавшей волны.

по-видимому, были забыты *). Позднее независимое и более полное исследование было выполнено Рабиновичем [105] (см. также [106]; исчерпывающий обзор исследований граничной волны дан в [107]). Теория Магги — Рабиновича в дальнейшем была развита Миамото и Вольфом [107] (см. также [109]).

Рассмотрим монохроматическую световую волну, распространяющуюся от точечного источника P_0 через отверстие в плоском непрозрачном экране. Как и раньше, предположим, что линейные размеры отверстия велики по сравнению с длиной волны, но малы по сравнению с расстояниями от P_0 и точки наблюдения P до экрана; допустим также, что углы падения света и углы дифракции малы. Тогда в приближениях, принятых в теории Кирхгофа (см. п. 8.3.2), имеем

$$U(P)=\frac{1}{4\pi}\iint\limits_{\mathcal{A}}\left\{\frac{e^{ikr}}{r}\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right)-\left(\frac{e^{iks}}{s}\right)\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right)\right\}dS, \qquad (1)$$

где $\mathcal{A}$ — дифракционное отверстие, а другие символы имеют такие же значения, как и раньше. Построим замкнутую поверхность, ограниченную отверстием $\mathcal{A}$, поверхностью усеченного конуса $\mathcal{B}$ с вершиной в P_0 и образующими, проходящими через край отверстия, и частью $\mathcal{C}$ большой сферы с центром в P (рис. 8.46). Если через R обозначить расстояние от P_0 до P, то получим строго из интегральной теоремы Кирхгофа

$$\frac{1}{4\pi}\iiint\limits_{\mathcal{A}+\mathcal{B}+\mathcal{C}}\left\{\frac{e^{ikr}}{r}\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right)-\frac{e^{iks}}{s}\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right)\right\}dS=\frac{e^{ikR}}{R} \quad \text{или} \quad 0, \qquad (2)$$

в зависимости от того, находится ли точка P внутри выбранной поверхности или вне ее. Теперь таким же путем, как и в п. 8.3.2, можно сделать вклад от $\mathcal{C}$

*) Работа Магги обсуждается также в [5, 102]. Экспериментальное подтверждение «существования» граничной волны можно найти в [103] (см. также [104]).

пренебрежимо малым, взяв радиус сферы достаточно большим. Тогда из (1) и (2) получим

$$U(P) = U^{(g)}(P) + U^{(d)}(P), \tag{3}$$

где

$$\left.\begin{aligned} &U^{(g)}(P) = \frac{e^{ikR}}{R}, && \text{если точка } P \text{ находится в прямом пучке,} \\ &U^{(g)}(P) = 0, && \text{если точка } P \text{ находится в области геометрической тени} \end{aligned}\right\} \tag{4}$$

и

$$U^{(d)}(P) = -\frac{1}{4\pi}\iint\limits_{\mathscr{B}} \left\{ \frac{e^{ikr}}{r} \frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \left(\frac{e^{iks}}{s}\right) \frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right) \right\} dS. \tag{5}$$

$U^{(g)}$ представляет возмущение, предсказанное геометрической оптикой, поэтому $U^{(d)}$ должно представлять влияние дифракции. Теперь покажем, что $U^{(d)}$ можно преобразовать в криволинейный интеграл по краю отверстия.

Отметим сначала, что сферы $r = \text{const}$ пересекают под прямым углом усеченный конус $\mathscr{B}$. Поэтому на $\mathscr{B}$

$$\frac{\partial}{\partial n}\left(\frac{e^{ikr}}{r}\right) = 0. \tag{6}$$

Кроме того,

$$\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) = \frac{\partial}{\partial s}\left(\frac{e^{iks}}{s}\right)\cos(n, s) = \left(\frac{ik}{s} - \frac{1}{s^2}\right) e^{iks}\cos(n, s). \tag{7}$$

Следовательно, (5) сводится к

$$U^{(d)}(P) = -\frac{1}{4\pi}\iint\limits_{\mathscr{B}} \frac{\exp[ik(r+s)]}{rs}\left(ik - \frac{1}{s}\right)\cos(n, s)\, dS. \tag{8}$$

Можно взять в качестве элемента dS площадь $ABB'A'$, ограниченную участками двух соседних образующих и дугами окружностей, по которым

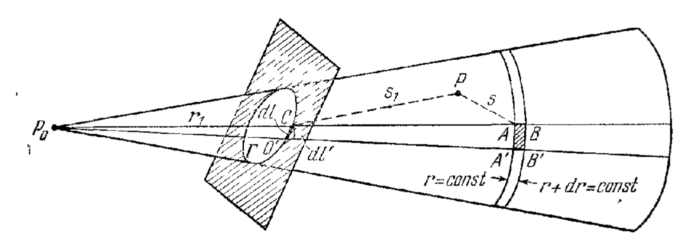


Рис. 8.47. К выводу выражения для граничной дифрагировавшей волны. $P_0A = r$, $P_0Q = r_1$, $PA = s$, $PQ = s_1$.

сферы $r = \text{const}$ и $r + dr = \text{const}$ пересекают конус (рис. 8.47). Если $d\varphi$ — угол между выбранными двумя образующими, то

$$dS = r\, dr\, d\varphi. \tag{9}$$

Пусть Q и Q' — точки пересечения двух образующих с краем Γ отверстия, и пусть dl — длина элемента Γ между этими двумя точками. Если dl' — соответствующий элемент дуги окружности, по которой сфера радиуса $r_1 = P_0Q$ пересекает конус, то

$$dl' = r_1\, d\varphi = dl\cos(dl, dl') = dl\sin(r_1, dl). \tag{10}$$

Из (9) и (10) имеем

$$dS=\frac{r}{r_1}\sin(r_1,\ dl)\,dr\,dl. \tag{11}$$

Далее, беря проекцию на нормаль к конусу в A и Q (нормали в этих точках параллельны), находим

$$s\cos(n,\ s)=s_1\cos(n,\ s_1). \tag{12}$$

Подставляя (11) и (12) в (8), получим

$$U^{(d)}(P)=-\frac{1}{4\pi}\int\int_{\mathscr{B}}\frac{\exp[ik(r+s)]}{rs}\left(ik-\frac{1}{s}\right)\frac{s_1}{s}\cos(n,\ s_1)\frac{r}{r_1}\sin(r_1,\ dl)\,dr\,dl=$$

$$=-\frac{1}{4\pi}\int_{\Gamma}dl\frac{s_1}{r_1}\cos(n,\ s_1)\sin(r_1,\ dl)\int_{r_1}^{\infty}\exp[ik(r+s)]\left(\frac{ik}{s^2}-\frac{1}{s^3}\right)dr. \tag{13}$$

Далее покажем, что подынтегральное выражение второго интеграла в (13) представляет собой полный дифференциал, т. е.

$$\exp[ik(r+s)]\left[\frac{ik}{s^2}-\frac{1}{s^3}\right]=\frac{d}{dr}\left\{\frac{\exp[ik(r+s)]}{s[s+r-r_1+s_1\cos(s_1,\ r_1)]}\right\}. \tag{14}$$

Выполнив дифференцирование в правой части (14), получим

$$\frac{d}{dr}\left\{\frac{\exp[ik(r+s)]}{s[s+r-r_1+s_1\cos(s_1,\ r_1)]}\right\}=\frac{\exp[ik(r+s)]}{s[s+r-r_1+s_1\cos(s_1,\ r_1)]}\times$$

$$\times\left\{ik\left(1+\frac{ds}{dr}\right)-\frac{1}{s}\frac{ds}{dr}-\frac{1}{s+r-r_1+s_1\cos(s_1,\ r_1)}\left[1+\frac{ds}{dr}\right]\right\}. \tag{15}$$

Теперь из треугольника APQ имеем

$$s^2=s_1^2+(r-r_1)^2+2s_1(r-r_1)\cos(s_1,\ r_1); \tag{16}$$

отсюда, дифференцируя по r и считая r_1 и s_1 постоянными, находим

$$s\frac{ds}{dr}=r-r_1+s_1\cos(s_1,\ r_1). \tag{17}$$

Подставляя это в (15), получим тождество (14). Следовательно,

$$\int_{r_1}^{\infty}\exp[ik(r+s)]\left(\frac{ik}{s^2}-\frac{1}{s^3}\right)dr=$$

$$=\left[\frac{\exp[ik(r+s)]}{s[s+r-r_1+s_1\cos(s_1,\ r_1)]}\right]_{r_1}^{\infty}=-\frac{\exp[ik(r_1+s_1)]}{s_1^2[1+\cos(s_1,\ r_1)]}, \tag{18}$$

и окончательно (13) принимает вид

$$U^{(d)}(P)=\frac{1}{4\pi}\oint_{\Gamma}\frac{\exp[ik(r_1+s_1)]}{r_1s_1}\frac{\cos(n,\ s_1)}{[1+\cos(s_1,\ r_1)]}\sin(r_1,\ dl)\,dl. \tag{19}$$

Это выражение совместно с (3) и (4) называется *«представлением Рабиносича»* дифракционного интеграла Кирхгофа и может рассматриваться как математическая формулировка теории Юнга. В ней дифракция рассматривается как взаимодействие падающей волны, распространяющейся по законам геометрической оптики, и граничной дифрагированной волны, которую можно считать возникшей вследствие рассеяния падающего света на границе отверстия.

Поскольку U — непрерывная функция положения, из соотношения (4) следует, что граничная волна $U^{(d)}$ испытывает разрыв на границе геометрической тени, который компенсирует разрыв «геометрической волны» $U^{(g)}$. Разрыв в $U^{(d)}$ обусловлен наличием в знаменателе (19) множителя $[1+\cos(s_1,\ r_1)]$.

Воспользовавшись аргументацией, совпадающей с доводами, выдвинутыми в связи с классификацией дифракционных явлений (см. стр. 355) и иллюстри-

рующими физический смысл принципа стационарной фазы (см. приложение 3), находим, что только те точки в области интегрирования вносят существенный вклад в $U^{(d)}$, для которых в подынтегральном выражении фаза постоянна, т. е.

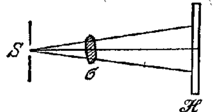
$$\frac{d}{dl}\left[k\left(r_1+s_1\right)\right]=0. \tag{20}$$

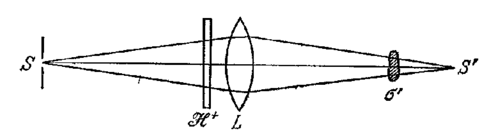
Это соотношение можно также записать в виде «закона отражения»

$$\cos\left(r_1,\ dl\right)=-\cos\left(s_1,\ dl\right). \tag{21}$$

§ 8.10. Метод Габора получения изображения восстановлением волновых фронтов *)

Исследуя пути, позволяющие улучшить разрешающую силу электронного микроскопа, Габор [110] создал новый двуступенчатый метод получения оптического изображения. На первой ступени объект освещается когерентной электронной или световой волной. Предполагается, что значительная часть волны проникает в объект без возмущения **). Дифракционная картина, образовавшаяся в результате интерференции вторичных волн, возникающих в присутствии объекта, с сильной опорной волной (когерентный фон), зарегистрированная на фотопластинке, называется *голограммой*. Если такую пластинку, надлежащим образом обработанную, поместить на место объекта и осветить только когерентной опорной волной, то прошедшая сквозь эту пластинку волна будет нести информацию об исходном объекте, которую можно извлечь из фотографии оптическими методами. Для «восстановления» объекта по этой «замещающей» волне необходимо только направить последнюю в подходящую систему, образующую изображение, и тогда изображение появится в плоскости, сопряженной с плоскостью, в которой находился объект. Здесь мы только коснемся оптических принципов этого метода***).





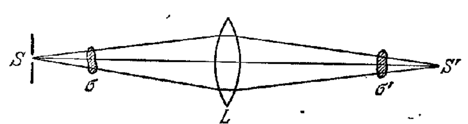


Рис. 8.48. Метод Габора получения изображения путем восстановления волновых фронтов.

8.10.1. Изготовление позитивной голограммы. Рассмотрим монохроматическую волну, идущую от небольшого источника S и падающую на полупрозрачный объект σ (рис. 8.48, *а*). Пусть $\mathscr{H}$ — экран, находящийся на некотором

) Подробное изложение теории голографии дается в работах [120—122*]. (*Прим. ред.*)

**) Это означает, что если волну представить в виде суммы падающей и вторичной дифрагировавшей волн, то рассеянием вторичных волн можно пренебречь. Указанное пренебрежение известно как первое борновское приближение и находит широкое применение в теории рассеяния рентгеновских лучей и электронов. Границы применимости этого приближения в электронной микроскопии и интерферометрии рассмотрены в [111].

***) Для детального изучения этого метода и его вариаций, а также для ознакомления с возможными применениями голографической техники см., например [112].

расстоянии за объектом, и пусть $U=Ae^{i\psi}$ — комплексное возмущение в произвольной точке $\mathscr{H}$; здесь A — амплитуда (вещественная) и ψ — фаза возмущения. Будем считать U суммой двух членов, а именно

$$U=U^{(i)}+U^{(s)}=\exp(i\psi_i)\left[A^{(i)}+A^{(s)}\exp\left[i(\psi_s-\psi_i)\right]\right]. \tag{1}$$

Здесь $U^{(i)}=A^{(i)}\exp(i\psi_i)$ — *падающая волна* (или *когерентный фон*); это поле, которое имелось бы на $\mathscr{H}$ в отсутствие объекта. Другой член $U^{(s)}=A^{(s)}\exp(i\psi_s)$ представляет *вторичную*, или *дифрагировавшую волну*; именно она несет информацию относительно объекта. Согласно (1) амплитуду полного возмущения U можно выразить через амплитуды и фазы этих волн в виде

$$A=\sqrt{UU^*}=\sqrt{(A^{(i)})^2+(A^{(s)})^2+2A^{(i)}A^{(s)}\cos(\psi_s-\psi_i)}. \tag{2}$$

Как обычно, мы опустили временной гармонический множитель $\exp(-i\omega t)$ и, следовательно, подразумеваем, что вторичные волны, исходящие от объекта, имеют частоту падающей волны. Такой объект называется *рассеивателем Рэлея*; почти все нефлуоресцирующие объекты можно с хорошим приближением отнести к этому типу.

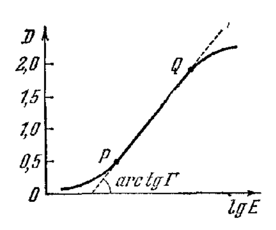


Рис. 8.49. Кривая Хертера — Дриффельда (характеристическая кривая).

Предположим теперь, что в плоскость $\mathscr{H}$ помещена фотографическая пластинка. Пусть α — *коэффициент пропускания* пластинки, определяемый аналогично функции пропускания F п. 8.6.1 как отношение комплексной амплитуды волны, прошедшей сквозь пластинку, к комплексной амплитуде волны, падающей на нее. Соответствующий коэффициент пропускания для интенсивности равен $\tau=\alpha\alpha^*$. Величина D, определяемая соотношением

$$D=-\lg\tau=-\lg\alpha\alpha^*, \tag{3}$$

называется оптической *плотностью* (или почернением) пластинки. Произведение E интенсивности $I=A^2$ света, падающего на пластинку, и времени выдержки t, т. е.

$$E=It, \tag{4}$$

называется просто *экспозицией* (или световой суммой). График зависимости D от $\lg E$ называется *кривой Хертера — Дриффельда* (характеристической кривой); типичная форма такой кривой показана на рис. 8.49. В области между точками P и Q кривая имеет вид практически прямой линии, и если ее наклон равен Γ, то, очевидно, для почернения негатива имеем

$$D=D_0+\Gamma\lg\frac{E}{E_0}, \tag{5}$$

где D_0 и E_0 — постоянные. Используя (3), получим

$$\tau=\tau_0\left(\frac{E}{E_0}\right)^{-\Gamma}. \tag{6}$$

При поглощении без изменения фазы α — вещественное число, равное квадратному корню из коэффициента пропускания для интенсивности, т. е. $\sqrt{\tau}$. Следовательно, в этом случае амплитудный коэффициент пропускания α_n «негативной голограммы» определяется соотношением вида

$$\alpha_n=(K_nA)^{-\Gamma_n}, \tag{7}$$

где K_n пропорционально квадратному корню из времени экспозиции.

Предположим теперь, что изготовлен позитивный отпечаток негативной голограммы. Тогда амплитудный коэффициент пропускания α_p позитива будет

равен

$$\alpha_p = [K_p (K_n A)^{-\Gamma_n}]^{-\Gamma_p} = KA^{\Gamma}, \tag{8}$$

где $\Gamma = \Gamma_n \Gamma_p$ — «общая гамма» негативно-позитивного процесса, а $K = K_p^{-\Gamma_n} K_n^{\Gamma}$.

8.10.2. Восстановление. В процессе восстановления (см. рис. 8.48, *б*) позитивная голограмма ($\mathscr{H}^+$), у которой амплитудный коэффициент пропускания определяется (8), освещается одним только когерентным фоном $U^{(i)}$. Когерентный фон осуществляется просто удалением объекта или иным способом с сохранением геометрии первоначальной схемы. Замещающая волна U', прошедшая сквозь пластинку, выражается, согласно (2) и (8), соотношением

$$U' = \alpha_p U^{(i)} = KA^{(i)} \exp(i\psi_i) \left[(A^{(i)})^2 + (A^{(s)})^2 + 2A^{(i)}A^{(s)} \cos(\psi_s - \psi_i)\right]^{\frac{1}{2}\Gamma}. \tag{9}$$

Если выбрать $\Gamma = 2$, то

$$U' = K (A^{(i)})^2 \exp(i\psi_i) \times$$
$$\times \left[A^{(i)} + \frac{(A^{(s)})^2}{A^{(i)}} + A^{(s)} \exp[i(\psi_s - \psi_i)] + A^{(s)} \exp[-i(\psi_s - \psi_i)]\right]. \tag{10}$$

Сравнивая (10) и (1), можно показать, что при постоянном $A^{(i)}$, т. е. при однородном когерентном фоне, замещающая волна U' содержит компоненту, называемую *восстановленной волной*, пропорциональную U (первый и третий члены в (10)). Остаток (10) состоит из двух членов. У одного из них фаза такая же, как у опорной волны, а амплитуда в $(A^{(s)})^2/A^{(i)}$ раз больше. Этот член можно сделать очень малым, применяя достаточно сильный когерентный фон *). У другого члена амплитуда $(K(A^{(i)})^2 A^{(s)})$ такая же, как и у восстановленной волны, а сдвиг фазы имеет знак, противоположный знаку фазы опорной волны. Этот член представляет *сопряженную волну*; можно считать, что она обусловлена воображаемым объектом, аналогичным истинному, но находящимся в другой плоскости.

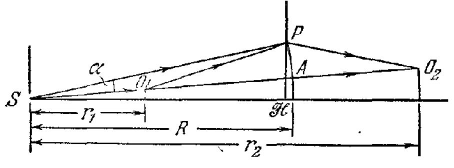


Рис. 8.50. Схема, иллюстрирующая положение сопряженного объекта.

Для того чтобы показать это, вернемся ненадолго к схеме, приведенной на рис. 8.48, *а*; обозначим через O_1 какую-нибудь точку объекта σ, а через P — какую-нибудь точку плоскости $\mathscr{H}$ (рис. 8.50). Можно допустить, что при освещении объекта точечным источником S свет приходит в точку P двумя путями, а именно по прямому пути SP (соответствующему опорной волне) и вдоль пути дифрагировавшего вторичного луча O_1P (соответствующему вторичной волне). Предположим сначала, что при дифракции на O_1 не происходит сдвига фаз. Тогда дифрагировавший свет в P отстает от прямого света, и разность хода равна $O_1P - O_1A$, где A — точка пересечения линии SO_1 со сферой радиуса SP и с центром в S. Сопряженная волна опережает прямую волну на такую же величину, на которую вторичная дифрагировавшая волна отстает от нее, так что сопряженная волна догоняет прямую в точке O_2 на продолжении линии SO_1, где

$$O_1P - O_1A = PO_2 - AO_2. \tag{11}$$

Пусть r_1, R и r_2 — расстояния от точек O_1, A и O_2 до S, а α — угол между SP

*) Сильный когерентный фон использован также в методе, предложенном Цернике [113] для выявления слабых вторичных интерференционных полос.

и линией SO_1AO_2 (см. рис. 8.50); тогда уравнение (11) можно переписать в виде

$$\sqrt{(R\cos\alpha-r_1)^2+R^2\sin^2\alpha}-(R-r_1)=\sqrt{(r_2-R\cos\alpha)^2+R\sin^2\alpha}-(r_2-R). \tag{12}$$

Разлагая обе части (12) в ряд по степеням α и оставляя только первые члены, получим

$$\frac{1}{r_1}+\frac{1}{r_2}=\frac{2}{R}. \tag{13}$$

Если бы голограмма была изогнута по сфере радиуса R, то тогда последнее соотношение осталось бы справедливым для любой точки на ней. Следовательно, в таком случае можно считать, что сопряженная волна обусловлена мнимым объектом, который в свою очередь является изображением первоначального объекта в сферическом зеркале радиуса R с центром в S (см. (4.4 16)). Очевидно, что этот результат как хорошее приближение справедлив и в случае плоской голограммы, если лучи составляют с осью небольшие углы Если, кроме того, r_1/R мало по сравнению с единицей, то, согласно (13), $r_2\approx-r_1$, и значит, сопряженный и истинный объекты располагаются симметрично относительно источника S.

До сих пор предполагалось, что дифракция на объекте не изменяет фазы падающей волны. Если же фаза изменяется, то предыдущее рассмотрение, касающееся положения сопряженного объекта, остается в силе при условии, что сопряженный объект вызывает такое же изменение фазы, как и истинный, но имеет противоположный знак.

Возвращаясь к уравнению (10), мы видим, что если когерентный фон однороден и достаточно силен по сравнению с рассеянной волной, то восстановленная волна U' оказывается точно такой же, как и первоначальная, за исключением вклада, который можно рассматривать как результат действия сопряженного объекта. Следовательно, *помещая линзу L позади позитивной голограммы, освещенной лишь сильным когерентным фоном* (см. рис. 8.48, *б*), *мы получим изображение σ' первоначального объекта в сопряженной плоскости, в которой лежит σ*. Вообще говоря, это изображение искажается вкладом, обусловленным сопряженным объектом. Однако можно найти условия, при которых искажающий эффект мал. Грубо говоря, можно ожидать, что он будет невелик тогда, когда расстояние между изображениями первоначального и сопряженного объектов превышает определяемую (8.8.27) величину допуска на положение фокальной плоскости для пучка, образующего изображение.

Недавно Лейт и Упатникс [114] значительно улучшили технику восстановления волнового фронта, изменив способ наложения когерентного фона на пучок, прошедший сквозь объект. В их установке, в которой использовался свет лазера, опорный пучок направлялся с помощью призмы или зеркальной системы под углом на плоскость голограммы. Сопряженное изображение в этом случае образуется в другом направлении и, следовательно, высококачественное изображение восстанавливается без всяких помех.

Голограмма содержит трехмерную информацию. Это также было показано Лейтом и Упатниксом и иллюстрируется рис. IV (см. вклейку).

На рис. 8.48, *в* приведена эквивалентная схема одноступенчатого процесса образования изображения, а на рис. III (см. вклейку) показаны увеличенный объект, голограмма и восстановленное изображение, полученное в одном из первых экспериментов такого рода.

Эти результаты подтверждают наш вывод о том, что, используя *когерентный свет*, можно восстановить изображение объекта с высокой степенью точности только по зарегистрированному *распределению интенсивности* в какой-нибудь плоскости, находящейся позади объекта.

Для успешного применения этого метода совершенно не обязательно, чтобы волновые фронты от источника S были строго сферическими. Необходимо только, чтобы геометрическая форма волновых фронтов опорной волны в схемах, изображенных на рис. 8.48, *а* и *б*, была одинаковой. Не обязательно также применение того же источника света или излучения с теми же длинами волн. В этой связи очевидны перспективы применения данного метода в электронной микроскопии. Одним из основных факторов, ограничивающих разрешающую силу электронного микроскопа, служит сферическая аберрация объектива; как отмечал Габор, влияние сферической аберрации в принципе можно устранить, или, лучше сказать, скомпенсировать с помощью метода восстановления. Для получения голограммы следует воспользоваться электронным пучком, для восстановления — видимым светом. При этом надо *имитировать* сферическую аберрацию электронного объектива, используя световые волны со сферической аберрацией такой же величины *) (конечно, измененной в отношении $\lambda_{свет}/\lambda_{электрон}$). Хотя в те годы, к которым относится это изобретение, трудности технического характера помешали использовать метод Габора в электронной микроскопии, правильность его основных принципов была проверена в экспериментах со светом.

Восстановление волнового фронта как метод двуступенчатой фотографии имело важного предшественника. Речь идет здесь о методе оптического фурье-анализа, применяемом для восстановления кристаллической структуры по ее дифракционной картине. Этот метод был впервые предложен Бэршом [116] и независимо от него Брэггом [117] (см. также [118]), который назвал свое устройство рентгеновским микроскопом. При освещении кристалла параллельным пучком рентгеновских лучей угловое распределение интенсивности дифрагировавшего света, зарегистрированное на фотопластинке, пропорционально, по крайней мере для малых углов отклонения, квадрату модуля фурье-образа проекции распределения электронной плотности в кристалле, спроектированной на плоскость, расположенную под прямым углом к первоначальному пучку **). То обстоятельство, что в этом опыте регистрируются только интенсивности, тогда как фазы остаются физически неопределенными, вообще говоря, мешает прямому восстановлению. Недостающие величины следует получать из других фотографий, из предшествующего изучения химической структуры и предварительных работ. Обработка измерений сравнительно проста, если направление освещения совпадает с осью кристалла; тогда можно показать, что при строго когерентном освещении фазы всех дифрагировавших волн либо совпадают, либо противоположны, и, значит, каждый наблюдаемый дифракционный максимум соответствует только двум возможным фазам. Во всяком случае, это обстоятельство значительно сокращает число комбинаций, которые следует испытать.

Существуют, однако определенные кристаллы, у которых каждая ячейка содержит тяжелый атом; тогда амплитуды волн, дифрагировавших на решетке из тяжелых атомов, столь велики, что результирующие амплитуды могут иметь только один знак. Можно сказать, что в этом случае тяжелые атомы служат причиной возникновения когерентного фона. Следовательно, при исследовании таких редко встречающихся кристаллов требуется только извлечь квадратный корень из интенсивностей дифрагировавшего света для того, чтобы получить фурье-образ проекции распределения плотности на плоскость, перпендикулярную той оси кристалла, которая параллельна направлению освещения.

Итак, распределение амплитуд, пропорциональных квадратному корню из интенсивности, получается из анализа фотографии (с общей гаммой, равной

---

*) Световые волны с заданной величиной сферической аберрации можно получить с помощью соответственно выбранных асферических поверхностей.

**) См. любой учебник по теории дифракции рентгеновских лучей, например [35, 36].

единице, см. уравнение (8)) дифракционной картины. В рентгеновском микроскопе на эту фотографию падает плоская волна и свет, прошедший сквозь фотографию и дифрагировавший на ней, фокусируется линзой. Согласно теории дифракции Фраунгофера (см. § 8.3) распределение амплитуд в плоскости изображения есть фурье-образ их распределения в плоскости пластинки и, следовательно, оно получается из первоначального распределения плотности двумя последовательными фурье-преобразованиями; из теоремы Фурье следует, что распределение амплитуд в фокальной плоскости есть точное изображение исходного (в общем случае комплексного) распределения плотности. Конечно, предполагается, что отверстие линзы достаточно велико и пропускает все дифрагировавшие лучи, несущие заметное количество энергии, и что удовлетворено требование синфазности волн всех длин. На самом же деле последнее условие выполняется редко. В методе восстановления волнового фронта искусственно добавляется когерентный фон. Однако не существует условий, обеспечивающих его симметрию или даже периодичность. Кроме того, этот метод не годится для рентгеновского анализа, так как там практически невозможно получить строго когерентный фон. Тем не менее он может найти успешное применение в будущих электронографических исследованиях.

ЛИТЕРАТУРА

1. C. F. Meyer, The Diffraction of Light, X-rays, and Material Particles, Univ. Press, Chicago, 1934.
2. A. Schuster, Phil. Mag. **31**, 77 (1891).
3. G. Kirchhoff, Berl. Ber., 641 (1882), Ann. d. Physik **18**, 663 (1883); Ges. Abh. Nachtr. 22.
4. F. Kottler, Ann. d. Physik **71**, 457 (1923); **72**, 320 (1923).
5. B. B. Baker, E. T. Copson, The Mathematical Theory of Huygens Principle, Clarendon Press. Oxford 2nd ed., 1950.
6. H. Helmholtz, J. Math. **57**, 7 (1859).
7. F. Pockels, Über die Partielle Differentialgleichung $(\nabla^2+k^2)\ U=0$ Teubner, Leipzig, 1891.
8. A. Sommerfeld, Optics, Acad. Press, N. Y. 1954, p. 199. (А. Зоммерфельд, Оптика, ИЛ 1953.)
9. M. Born, Optik, Springer, Berlin, 1933, p. 149. (М. Борн, Оптика, ГНТИУ, Харьков—Киев, 1937.)
10. G. G. Stokes, Trans. Cambr. Phil. Soc. **9**, 1 (1849).
11. A. Babinet, C. R. Acad. Sci. **4**, 638 (1837).
12. H. Poincaré, Théorie mathématique de la lumière, Paris, 1892, Vol. II, p. 187.
13. G. Toraldo di Francia, Atti Fond Giorgio Ronchi **11**, § 6 (1956).
14. E. Wolf, E. W. Marchand, J. Opt. Soc. Amer. **5B**, 1712 (1966).
14a. F Kottler, Ann. d. Physik 70, 405 (1923).
15. F. Kottler, в сб.: «Progr. in Optics», vol. 4, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1964.
16. C. J. Boukwkamp, в сб. «Rep. Progr. Phys.», London, 1954, Vol. 17, p. 35.
17. S. Silver, J. Opt. Soc. Amer. **52**, 131 (1962).
18. I. N. Sneddon, Fourier Transforms, McGraw-Hill, N. Y., 1951, pp. 25, 44.
19. O. Theimer, G. D. Wassermann, E. Wolf, Proc. Roy. Soc. **A212**, 426 (1952).
20. E. Wolf, Proc. Phys. Soc. **74**, 269 (1959).
21. M. Planck, Ann. d. Physik **1**, 61 (1900)
22. M. Born, P. Jordan, Z. f. Phys. **33**, 479 (1925).
23. W. Gröbner, N. Hofreiter, Integraltafel, Wien, 1950, Vol. II, p. 333.
24. E. Jahnke, F. Emde, Tables of Functions with Formulae and Curves, Teubner, Leipzig, 1933. (Е. Янке, Ф. Эмде, Специальные функции, формулы, графики и таблицы, «Наука», 1964.)
25. G. N. Watson, A. Treatise on the Theory of Bessel Functions, Cambr. Univ. Press, 1944. (Г. Н. Ватсон, Теория бесселевых функций, ИЛ, 1949.)
26. E. T. Whittaker, G. N. Watson, A Course of Modern Analysis, Cambr. Univ. Press, 4th ed., 1952. (Э. Т. Уиттекер, Г. Н. Ватсон, Курс современного анализа, Физматгиз, 1962.)
27. G. B. Airy, Trans. Cambr. Phil. Soc. **5**, 283 (1835).

28. В. Игнатовский, Труды Оптического института, Петроград, 1, IV (1919); В. А. Фок, Труды Оптического института 3, 24 (1924); A. Boivin, E. Wolf, Phys. Rev. 138B, 1561 (1965), A. Boivin, J. Dow. E. Wolf, J. Opt. Soc. Amer., 57, 1171 (1967).
29 Rayleigh, Phil. Mag. 11, 214 (1881).
30. Rayleigh, Scientific Papers, Cambr. Univ. Press, 1899, Vol. 1, p. 513.
31. J. Scheiner, S. Hirayma, Abh. d. Königl. Akad. Wissensch., Berlin, 1894, Anhang I.
32. J. V. Kathavate, Proc. Indian Acad. Sci. 21, 177—210 (1945).
33. M. Laue, Berl. Ber., 1144 (1914).
34. J. A. Prins, Naturwiss. 19, 435 (1931).
35. M. Laue, Röntgenstrahl-Interferenzen, Akad. Verlagsgesellschaft, 2 Aufl, 1948.
36. В сб. «The Crystalline State», ed. L. Bragg. Bell a. Sons, London, Vol. I, W. L. Bragg, A. General Survey, 1933, Vol. II, W. R. James. The Optical Principles of the Diffraction of X-rays, 1948; Vol. III; H. Lipson, W. Cochran, The Determination of Crystal Structures, 1953.
37. W. Friedrich, P. Knipping. M. Laue, Münchener Sitzungsber., 1912, p. 303; Ann. d. Physik 61, 971 (1913).
38. G. R. Harrison, J. Opt. Soc. Amer. 39, 413 (1949).
39. D. Rittenhouse, Trans. Amer. Phil. Soc. 2, 201 (1786).
40. T. D. Cope, J. Franklin Inst. 214, 99 (1932).
41. J. Fraunhofer, Denkschr. Akad. Wiss. München 8, 1 (1821—1822). Ann. d. Physik 74, 337 (1823).
42. H. A. Rowland, Phil. Mag. 13, 469 (1882); Nature 26, 211 (1882). Phil. Mag. 16, 297 (1883).
43. R. W. Wood, Phil. Mag. 20, 770 (1910); Phil. Mag. 23, 310 (1912); A. Trowbridge, R. W. Wood, Phil. Mag 20, 886, 898 (1910).
44. R. W. Wood, Nature 140, 723 (1937); J. Opt. Soc. Amer. 34, 509 (1944); H. Babcock, J. Opt. Soc. Amer. 34, 1 (1944).
45. G. R. Harrison, J. Opt. Soc. Amer. 39, 522 (1949).
46. G. W. Stroke, в сб.: «Progr. in Optics», vol. 2, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1963, p. 1.
47. T. Thorp, British Patent № 11460, 1899.
48. R. J. Wallace, Astrophys, J. 22, 123 (1905); 23, 96 (1906).
49. T. Merton, Proc. Roy. Soc. A201, 187 (1950).
50. A. A. Michelson, Astrophys. J. 8, 37 (1898); Proc. Amer. Acad. Arts Sci. 35, 109 (1899).
51. W. E. Williams, British Patent, № 3 12534 1926 Proc. Opt. Conv. 2, 982 (1926). Proc. Soc. 45, 699 (1933).
52. F. L. O Wadsworth, Astrophys. J. 3,, 54 (1896).
53. W. T. Welford, в сб.: «Progr. in Optics», vol. 4, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1965, p.241.
54. G. C. Steward, The Symmetrical Optical System, Cambr. Univ. Press, 1928, p. 89.
55. R. Straubel, P. Zeeman Verh., Martinus Nijhoff , 1935, p. 302; R. K. Luneburg, Mathematical Theory of Optics, Brown Univ., Providence, R I. (1944), p 391, печатное издание Univ. of California Press, Berkeley a Los Angeles, 1964, § 50; H. Osterberg, J E. Wilkins, J. Opt. Soc. Amer. 39, 553 (1949), G. Lansraux, Rev. d'Opt. 32, 475 (1953).
56. E. Wolf, в сб. «Rep. Progr. Phys.» Phys. Soc., London, 1951, Vol. 14, p. 109.
57. G. Toraldo di Francia, Suppl. Nuovo Cimento 9, 426 (1952).
58. P. Jacquinot, B. Roizen-Dossier, в сб.: «Progr. in Optics», vol. 3; ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1964, p. 29.
59. E. Abbe, Archiv f. Mikroskopische Anat. 9, 413(1873).
60. E. Abbe, Gesammelte Abhandlungen, G Fischer, Jena, 1904, I, p. 45; O. Lummer, F. Reiche. Die Lehre von der Bildtstehung im Mikoskop von Ernst Abbe, Vieweg, Braunschweig, 1910.
61. A. B. Porter, Phil. Mag. 11, 156 (1906).
62 Rayleigh, Phil. Mag 42, 167 (1896).
63. Rayleigh, Scientific Papers. Cambr Univ. Press; 1903, Vol. 4, p. 235.
64. R. Courant, D. Hilbert. Methods of Mathematical Physics, Intersci. Publ., N. Y., 1953, vol. 1, p. 79. (Р. Курант, Д. Гильберт, Методы математической физики, Гостехиздат. 1945.)
65. M. Francon, Le contraste de phase en optique et en microscopie. Rev. d'Opt. 1950; A. H Bennett, H. Jupnik, H. Osterberg, O. W. Richards, Phase Microscopy, J. Wiley a. Sons, N. Y., 1952.
66. H. H. Hopkins, в кн. M. Francon, Le contraste de phase et le contraste par interférences, Rev. d'Opt., 1952, p. 142.

67. F. Zernike, Z. Tech. Phys. **16**, 454 (1935); Phys. Z. **36**, 848 (1935); Physica **9**, 686, 974 (1942).
68. J. Picht, Z. f. Instrkde **56**, 481 (1936); Z. f. Instrkde **58**, 1 (1938).
69. F. D. Kahn, Proc. Phys. Soc., **B68**, 1073 (1955).
70. A. H. Bennett, H. Jupnik, H. Osterberg, O. W. Richards, Trans. Amer. Microscop. Soc. **65**, 126 (1946).
71. A. H. Bennett, H. Jupnik, H. Osterberg, O. W. Richards. Trans. Amer. Microscop. Soc. **65**, 119 (1946).
72. F. Zernike, Mo. Not Roy. Astr. Soc. **94**, 382—383 (1934).
73. British Association Report, Oxford, 1926, pp. 273—275. T. Pearcey, Tables of Fresnel Integrals to Six Decimal Places, Cambr. Univ. Press, 1956.
74. R. Courant, Differential and Integral Calculus, Blackie a. Sons Ltd, 2nd ed., 1942, vol. 1, p. 496. (Р. Курант, Курс, дифференциального и интегрального исчисления, «Наука», 1967.)
75. A. Cornu, J. de Phys. **3**, 5, 44 (1874).
76. E. Lommel, Abh. Bayer. Akad **15**,, Abth. 2, 233 (1885)
77. E. Lommel, Abh. Bayer. Akad. **15**, Abth. 3, 531 (1886).
78. H. Struve, Mém. de l'Acad. de St. Petersbourgh (7), **34**, 1 (1886).
79. K. Schwarzschild, Sitzb. München. Akad. Wiss., Math-Phys., Kl. **28**, 271 (1898).
80. P. Debye, Ann. d. Physik **30**, 775 (1909).
81. A. Gray, G B. Matthews, T M. MacRobert, A Treatise on Bessel Functions, Macmillan, London, 2nd ed., 1922, Ch. XIV.
82. J. Walker, The Analytical Theory of Light, Cambr. Univ. Press, 1904, p. 396.
83. M. E. Hufford, H. T. Davis. Phys. Rev. **33**, 589 (1926).
84. C. A. Taylor, B. J. Thompson, J. Opt. Soc. Amer. **48**, 844 (1958).
85. M. P. Bachynski, G. Bekefi, J. Opt. Soc. Amer. **47**, 428 (1957).
86. P. A. Mathews, A. L. Cullen, Proc. IEE, Pt. C. **103**,, 449 (1956).
87. E. H. Linfoot, E. Wolf, Proc. Phys. Soc. B69, 823 (1956).
88. M. Berek, Z. Phys. **40**, 421 (1926).
89. F. Zernike, B. R. A. Nijboer, в сб. «La Théorie des Images Optiques», Rev. d'Opt. Paris, 1949, p. 227.
89a. A. Boivin, E. Wolf, Phys. Rev. **138B**, 1561 (1965); A. Boivin, J. Dow, E. Wolf, J. Opt. Soc. Amer. **57**, 1171 (1967).
90. E. H. Linfoot, E. Wolf, Proc. Phys. Soc. **B66**, 145 (1953).
91. A. Boivin, J. Opt. Soc. Amer, **42**, 60 (1952).
92. H. D. Taylor, Mo Not. Roy. Astr. Soc. **54**, 67 (1894).
93. E. Wolf, Proc. Roy. Soc. **A204**, 533 (1951).
94. J. Focke, Optica Acta **3**, 161 (1956).
95. G. W. Farnell, Canad. J. Phys. **35**, 780 (1957).
96. L. G. Gouy, C. R. Acad. Sci. **110**, 1251 (1890); Ann. Chim. (Phys.) (6), **24**, 145 (1891).
97. F. Reiche, Ann. d. Physik **29**, 56, 401 (1909).
98. A. Rubinowicz, Phys. Rev. **54**, 931 (1938); C. J. Bouwkamp. Physica **7**, 485 (1940).
99. Farnell, Canad. J. Phys. **36**, 935 (1958).
100. T. Young, Phil. Trans. Roy. Soc. **20**, 26 (1802).
101. G. A. Maggi, Ann. di Matem. **16**, 21 (1888).
102. F. Kottler, Ann. d. Physik **70**, 413 (1923).
103. W. Wien, Inaug. Diss., Berlin, 1886; E. Maey, Ann. d. Physik **49**, 69 (1893); А. Калашников, Ж. Русск. физ.-хим. об-ва, Физ. часть **44**, 133 (1912).
104. S. Banerji, Phil. Mag. **37**, 112 (1919); S. K. Mitra, Phil. Mag. **38**, 289 (1919).
105. A. Rubinowicz, Ann. d. Physik **53**, 257 (1917); **73** (1924); **81**, 153 (1926); Acta Phys. Polon. **12**, 225 (1953).
106. G. N. Ramachadran, Proc. Indian. Acad. Sci. **A21**, 165 (1945); L. C. Martin, Proc. Phys. Soc. **55**, 104 (1943); **362**, 713 (1949); Y. Y. Kathavate, Proc. Phys. Soc. **A21**, 177 (1945); R. S. Ingarden, Acta Phys. Polon. **14**, 77 (1955); O. Laporte, J. Meixner, Z. f. Phys. **153**, 129 (1958).
107. A. Rubinowicz, Die Beugungswelle in der Kirchhoffschen Theorie der Beugung, Polska Akad. Nauk, 1957.
108. K. Miyamoto E. Wolf, J. Opt. Soc. Amer. **52**, 615, 626 (1962); K. Miyamoto, Proc. Phys. Soc. **79**, 617 (1962).
109. E. W. Marchand, E. Wolf, J. Opt. Soc. Amer. **52**, 761 (1962); A. Rubinowicz, J. Opt. Soc. Amer. **52**, 717 (1962); Acta Phys. Polon. **21**, 61, 451 (1962); сб.: «Progr. in Optics». vol. 4, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y. 1965, p. 199.
110. D. Gabor, Nature, **161**, 777 (1948); Proc. Roy. Soc. **A197**, 454 (1949); Proc. Phys Soc. **B64**, 449 (1951). (Русские переводы последних статей опубликованы в [120*].)
111. D. Gabor. Rev. Phys. **28**, 260 (1956).

112. J. B. DeVelis, G. O. Reynolds, Theory and Applications of Holography, Reading Mass Addison Wesley Publ., Co., 1964; H. M. Smith, Principles of Holography, J. Wiley a. Sons, N. Y., 1969.
113. F. Zernike, Proc. Phys. Soc., **61**, 158 (1948)
114. E. N. Leith, Upatnieks, J. Opt. Soc. Amer. **52**, 1123 (1962); **53**, 1377 (1963); **54**, 1295 (1964).
115. E. N. Leith, Upatniekc, J. Opt. Soc. Amer. **54**, 1297 (1964).
116. H. Boersch, Z. techn. Phys **19**, 337 (1938).
117. W. L. Bragg, Nature **149**, 470 (1942).
118. M. J. Buerger, J. Appl. Phys. **21**, 909 (1950).
119*. Л. И. Мандельштам, Ann. d. Physik **35**, 881 (1911); **39**, 493 (1912); Полное собрание трудов, т. I, Изд. АН СССР, 1948, стр. 211, 226, 229; Д. С. Рождественский, ЖЭТФ **10**, 305 (1940).
120*. Дж. Строук, Введение в когерентную оптику и голографию, «Мир», 1967.
121*. Л. М. Сороко, Лекции по голографии. Препарат ОИЯИ № 2587.
122*. Ю. Н. Денисюк, Об отображении оптических свойств объекта в волновом поле рассеянного им излучения, ДАН СССР **144**, 1275—1278 (1962).

ГЛАВА 9

ДИФРАКЦИОННАЯ ТЕОРИЯ АБЕРРАЦИЙ

В гл. 5 мы изучали эффекты аберраций, пользуясь приближением геометрической оптики. Изображением считалась размазанная фигура, образованная точками пересечения геометрических лучей с плоскостью изображения. Поскольку геометрическая оптика дает хорошее приближение в предельном случае очень коротких волн, естественно ожидать, что геометрическая теория аберраций постепенно перестает быть справедливой при уменьшении величины аберрации. Например, в предельном случае идеально сферической сходящейся волны, выходящей из круглого отверстия, геометрическая оптика предсказывает бесконечную интенсивность в фокусе и нулевую интенсивность на всей остальной фокальной плоскости, тогда как на самом деле изображение (см. п. 8.5.2) состоит из яркого центрального пятна, окруженного темными и светлыми полосами (картина Эйри). Было показано также, что распределение света в непосредственной близости от фокальной плоскости значительно сложнее (см. рис. 8.39), чем следовало бы ожидать на основании предсказаний геометрической оптики. Поэтому эффекты аберраций необходимо исследовать на основе теории дифракции.

Первые работы в этой области принадлежат Рэлею [1]. В них наибольшее значение имеет формулировка критерия (рассмотренного в § 9.3), который в расширенной форме получил широкое применение для определения максимальной величины аберраций, допустимых в оптических приборах. Исследованиям в этой области посвящены работы многих ученых, изучавших эффекты различных аберраций *); отметим, в частности, наиболее крупные из таких работ [3—9].

Очень обширное исследование по дифракционной теории формирования изображения при наличии аберраций принадлежит Нижберу **) [10]; оно частично выполнено им совместно с Цернике ***) [13]. Эта работа посвящена эффектам малых аберраций, при которых отклонения волновых фронтов от сферической формы составляют доли длины волны. Ван Кампен [17—19] рассмотрел эффекты больших аберраций, пользуясь асимптотическими приближениями в теории дифракции; его исследование основано на формальном переходе к функциям двух переменных в методе стационарной фазы, впервые строго сформулированном Фокке (см. приложение 3).

Основная часть настоящей главы посвящена обзору теории Нижбера — Цернике и рассмотрению структуры дифракционных изображений, искаженных первичными аберрациями. В последнем параграфе (см. § 9.5) мы перейдем от точечных предметов к протяженным и исследуем изображения, получающиеся при когерентном и некогерентном освещении. Отображения при частично когерентном освещении будут рассмотрены в гл. 10.

---

*) Исторический обзор дифракционной теории аберраций написан Вольфом [2].

**) Основная часть этой работы опубликована также в [11, 12].

***) Обобщение этой теории на случай больших аберраций содержится в [14]. Отметим также работу [15], которая в основном посвящена экспериментальному изучению эффектов аберраций. В § 9.4 приведено несколько прекрасных фотографий, полученных Нинуисом.

В статье Бачинского и Бекефи, ссылка на которую имеется в гл. 8, а также в [16] содержится описание экспериментального исследования структуры области изображения при наличии аберраций в микроволновом диапазоне частот.

§ 9.1. Дифракционный интеграл при наличии аберраций

9.1.1. Дифракционный интеграл. Рассмотрим центрированную оптическую систему с точечным источником монохроматического света P_0 (рис. 9.1). Выберем декартову систему координат, начало которой находится в месте параксиального изображения P_1^* точки P_0, а ось z направлена вдоль CP_1^*, где C — центр выходного зрачка. Ось y лежит в меридиональной плоскости (плоскости, в которой лежит точка P_0 и ось системы). Обозначим через Y_0 и Y_1^* расстояния от точек P_0 и P_1^* до оси.

Как и в гл. 5, деформацию волновых фронтов в области выходного зрачка опишем функцией аберраций Φ. Пусть $\bar{Q}$ и Q — точки пересечения луча в пространстве изображения с волновым фронтом, проходящим через C, и опорной

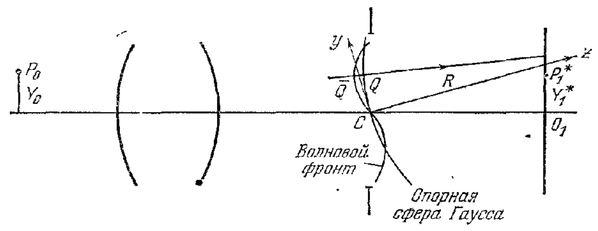


Рис. 9.1. Выбор системы отсчета и принятые обозначения.

сферой Гаусса соответственно. Если предположить, что показатель преломления среды в пространстве изображения равен единице, то Φ (на рис. 9.1 $\Phi > 0$) является расстоянием $\bar{Q}Q$, измеренным вдоль луча.

Пусть R — радиус опорной сферы Гаусса CP_1^*, s — расстояние между Q и произвольной точкой P, расположенной вблизи изображения. Возмущение в точке Q описывается величиной $A \exp[ik(\Phi - R)]/R$, где A/R — амплитуда возмущения в точке Q. Согласно принципу Гюйгенса — Френеля возмущение в точке P равно

$$U(P) = -\frac{i}{\lambda}\frac{A \exp(-ikR)}{R}\iint \frac{\exp[ik(\Phi + s)]}{s}\,dS, \qquad (1)$$

где интегрирование проводится по той части опорной сферы Гаусса, которая примерно закрывает выходной зрачок. При написании (1) предполагалось, что входящие в это выражение углы малы, так что можно пренебречь изменением коэффициента наклона лучей по опорной сфере Гаусса; кроме того, предполагалось, что амплитуда волны практически постоянна вдоль волнового фронта, т. е. коэффициент A можно вынести из-под знака интеграла.

Пусть (ξ, η, ζ) и (x, y, z) — координаты точек Q и P, и a — радиус выходного зрачка. Как и в § 8.8, где рассматривалась волна, свободная от аберраций ($\Phi \equiv 0$), положим

$$\left.\begin{aligned} \xi &= a\rho \sin\theta, & x &= r \sin\psi, \\ \eta &= a\rho \cos\theta, & y &= r \cos\psi, \end{aligned}\right\} \qquad (2)$$

тогда *), как и в (8.8.2) и (8.8.9), имеем

$$k(s - R) = -v\rho \cos(\theta - \psi) - \frac{1}{2}u\rho^2 + \left(\frac{R}{a}\right)^2 u, \qquad (3)$$

*) Здесь R соответствует величине f из § 8.8. Интересно отметить, что дифракционный интеграл можно снова представить в виде разложения по угловым гармоникам, если воспользоваться уравнениями (8.8.2) и (8.8.3) [20].

где u и v — «оптические координаты» точки P, т. е.

$$u=\frac{2\pi}{\lambda}\left(\frac{a}{R}\right)^2 z, \quad v=\frac{2\pi}{\lambda}\left(\frac{a}{R}\right)\sqrt{x^2+y^2}. \tag{4}$$

Величину Φ удобно рассматривать как функцию Y_1^*, ρ и θ

$$\Phi=\Phi(Y_1^*, \rho, \theta). \tag{5}$$

Элемент поверхности опорной сферы Гаусса dS равен $a^2\rho\, d\rho\, d\theta$, и если угол, который CP_1^* составляет с осью системы, мал, то областью интегрирования в (1) может служить $0\leqslant\rho\leqslant 1$, $0\leqslant\theta\leqslant 2\pi$. Кроме того, для точек наблюдения, находящихся вблизи изображения, величину s, стоящую в знаменателе подынтегрального выражения, можно заменить на R. Таким образом, соотношение (1) после подстановки в него (3) принимает вид

$$U(P)=U(u, v, \psi)=-\frac{i}{\lambda}\frac{Aa^2}{R^2}\exp\left[i\left(\frac{R}{a}\right)^2 u\right]\times$$
$$\times\int_0^1\int_0^{2\pi}\exp\left\{i\left[k\Phi(Y_1^*, \rho, \theta)-v\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]\right\}\rho\, d\rho\, d\theta, \tag{6}$$

и интенсивность в точке P равна

$$I(P)=|U(P)|^2=$$
$$=\left(\frac{Aa^2}{\lambda R^2}\right)^2\left|\int_0^1\int_0^{2\pi}\exp\left\{i\left[k\Phi(Y_1^*, \rho, \theta)-v\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]\right\}\rho\, d\rho\, d\theta\right|^2. \tag{7}$$

Удобно выразить $I(P)$ в виде доли интенсивности I^*, которая получалась бы в точке параксиального изображения P_1^* в отсутствие аберраций. Согласно (7)

$$I^*=\pi^2\left(\frac{Aa^2}{\lambda R^2}\right)^2; \tag{8}$$

тогда *нормированная интенсивность* *) запишется в виде

$$i(P)=\frac{I(P)}{I^*}=$$
$$=\frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi}\exp\left\{i\left[k\Phi(Y_1^*, \rho, \theta)-v\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]\right\}\rho\, d\rho\, d\theta\right|^2. \tag{9}$$

В отсутствие аберраций интенсивность максимальна в точке параксиального изображения. При наличии же аберраций максимум интенсивности расположен, в общем случае, в другой точке, которую можно назвать *дифракционным фокусом* **). Часто интерес представляет лишь максимальная интенсивность, получающаяся в определенной плоскости наблюдения; эта величина (если она нормирована, как и (9)) называется *интенсивностью Штреля* ***).

Из (9) можно сразу же получить несколько простых результатов, которые понадобятся позднее.

9.1.2. Теорема смещения. Изменение опорной сферы. Пусть Φ и Φ' — две функции аберраций, причем

$$\Phi'=\Phi+H\rho^2+K\rho\sin\theta+L\rho\cos\theta+M, \tag{10}$$

*) Символ i для нормированной интенсивности нельзя спутать с таким же обозначением для $\sqrt{-1}$, поскольку первый всегда пишется с аргументами, например $i(P)$, $i(u, v, \psi)$ и т. д.

**) Вообще говоря, конечно, может быть несколько дифракционных фокусов, однако если аберрации достаточно малы, имеется только один дифракционный фокус.

***) Это понятие ввел Штрель [21], который назвал его «Definitionshelligkeit». В английской литературе часто используется менее подходящий термин «definition» (определение).

где H, K, L, M — постоянные порядка λ. Далее пусть $i(u, v, \psi)$ и $i'(u, v, \psi)$ — соответствующие нормированные интенсивности. Тогда из (9) имеем

$$i(u, v, \psi)=\frac{1}{\pi^2}\left|\int\limits_0^1\int\limits_0^{2\pi}\exp\left[if(u, v, \psi; \rho, \theta)\right]\rho\,d\rho\,d\theta\right|^2, \tag{11}$$

где

$$f(u, v, \psi; \rho, \theta)=k\Phi-v\rho\cos(\theta-\psi)-{}^1/_2\,u\rho^2. \tag{12}$$

Аналогичное выражение справедливо и для i'. Согласно (10) последнее выражение можно представить в виде

$$\begin{aligned}f(u, v, \psi; \rho, \theta)&=k\Phi'-k\left[H\rho^2+K\rho\sin\theta+L\rho\cos\theta+M\right]-\\&-v\rho\cos(\theta-\psi)-{}^1/_2\,u\rho^2=k\Phi'-v'\rho\cos(\theta-\psi')-{}^1/_2\,u'\rho^2-kM=\\&=f'(u', v', \psi'; \rho, \theta)-kM,\end{aligned} \tag{13}$$

где

$$\left.\begin{aligned}u'=u+2kH,\qquad v'\sin\psi'&=v\sin\psi+kK,\\ v'\cos\psi'&=v\cos\psi+kL.\end{aligned}\right\} \tag{14}$$

В соответствии с (2) и (4) уравнения (14) описывают преобразование

$$z'=z+2\left(\frac{R}{a}\right)^2 H,\qquad x'=x+\left(\frac{R}{a}\right)K,\qquad y'=y+\left(\frac{R}{a}\right)L. \tag{15}$$

Из (11) и (13) вытекает, что

$$i(u, v, \psi)=i'(u', v', \psi'). \tag{16}$$

Таким образом, мы доказали следующую *теорему смещения. Добавление к функции аберраций члена* $H\rho^2+K\rho\sin\theta+L\rho\cos\theta+M$, где H, K, L, M — *постоянные порядка* λ, *не изменяет трехмерного распределения интенсивности света близ фокуса, а только смещает его как целое в соответствии с преобразованиями* (15), иными словами, происходит смещение на величину $2(R/a)^2H$ вдоль главного направления CP_1^* от выходного зрачка и на величины $(R/a)K$ и $(R/a)L$ вдоль положительных направлений осей x и y соответственно.

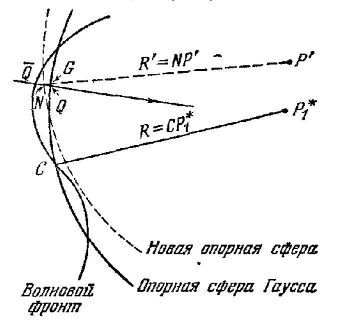


Рис. 9.2. Изменение опорной сферы.

Аддитивные члены в правой части (10) можно рассматривать как величины, характеризующие изменение опорной сферы Гаусса. Предположим, что мы выбрали новую опорную сферу с центром в точке $P'(x', y', z')$ вблизи изображения и радиусом R', причем ее расстояние до опорной сферы Гаусса не превышает нескольких длин волн. Пусть N — точка пересечения луча $\bar{Q}Q$ с этой новой опорной сферой Тогда функция аберрации Φ', отнесенная к новой сфере, имеет вид (см. рис. 9.2)

$$\Phi'=\bar{Q}N=\bar{Q}Q-NQ\approx\bar{Q}Q-NG, \tag{17}$$

где G — точка пересечения линии NP' с опорной сферой Гаусса, а показатель преломления среды в пространстве изображения считается, как и раньше, равным единице. Величина $\bar{Q}Q=\Phi$ есть волновая аберрация, отнесенная к опорной сфере Гаусса, а $NG=NP'-GP'=R'-s$, где s — расстояние от G до P'. Следовательно, (17) можно переписать в виде

$$\Phi'\approx\Phi+s-R'=\Phi+\frac{\lambda}{2\pi}\left[-v\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2+\left(\frac{R}{a}\right)^2u\right]+(R-R'). \tag{18}$$

где было использовано соотношение (3). Здесь u, v и ψ определяются (3) и (4), где координаты x, y, z нужно заменить на x', y', z'. Соотношение (18) можно представить в форме (10), если положить

$$\left.\begin{aligned} H &= -\frac{1}{2}\left(\frac{a}{R}\right)^2 z', \qquad & K &= -\left(\frac{a}{R}\right)x', \\ L &= -\left(\frac{a}{R}\right)y', & M &= z+R-R'. \end{aligned}\right\} \tag{19}$$

9.1.3. Соотношение между интенсивностью и средней деформацией волновых фронтов. Если аберрации достаточно малы, то интенсивность в центре опорной сферы можно выразить через среднеквадратичное значение волновой аберрации. Пусть Φ_p — волновая аберрация, отнесенная к опорной сфере с центром в точке P вблизи изображения. Тогда, согласно уравнениям (9) и (18), нормированную интенсивность в точке P можно записать следующим образом:

$$i(P) = \frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi} \exp(ik\Phi_p)\,\rho\,d\rho\,d\theta\right|^2 = \frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi}\left[1 + ik\Phi_p + \frac{1}{2}(ik\Phi_p)^2 + \ldots\right]\rho\,d\rho\,d\theta\right|^2. \tag{20}$$

Пусть $\overline{\Phi_p^n}$ — среднее значение n-й степени Φ_p, т. е.

$$\overline{\Phi_P^n} = \frac{\int_0^1\int_0^{2\pi}\Phi_P^n\,\rho\,d\rho\,d\theta}{\int_0^1\int_0^{2\pi}\rho\,d\rho\,d\theta} = \frac{1}{\pi}\int_0^1\int_0^{2\pi}\Phi_P^n\,\rho\,d\rho\,\rho\theta. \tag{21}$$

Если считать аберрации настолько малыми, что в (20) можно пренебречь третьей и более высокой степенями $k\Phi_p$, то выражение для интенсивности в точке P можно представить в форме

$$i(P) \approx \left[1 + ik\overline{\Phi}_p - \frac{1}{2}k^2\overline{\Phi_P^2}\right]^2 = 1 - \left(\frac{2\pi}{\lambda}\right)^2\left[\overline{\Phi_P^2} - (\overline{\Phi}_p)^2\right]. \tag{22}$$

Величина в скобках в правой части (22) является «среднеквадратичной деформацией» $(\Delta\Phi)^2$ волнового фронта, т. е.

$$(\Delta\Phi_p)^2 = \frac{\int_0^1\int_0^{2\pi}(\Phi_P - \overline{\Phi}_P)^2\,\rho\,d\rho\,d\theta}{\int_0^1\int_0^{2\pi}\rho\,d\rho\,d\theta} = \overline{\Phi_P^2} - (\overline{\Phi}_p)^2, \tag{23}$$

и, следовательно, (22) принимает вид

$$i(P) \approx 1 - \left(\frac{2\pi}{\lambda}\right)^2(\Delta\Phi_p)^2. \tag{24}$$

Отсюда следует, что в случае малых аберраций значение нормированной интенсивности в центре опорной сферы вблизи фокуса не зависит от природы аберраций и отличается от единицы, соответствующей идеальному случаю, на величину, пропорциональную среднеквадратичной деформации волнового фронта.

§ 9.2. Разложение функции аберрации

9.2.1. Круговые полиномы Цернике. При изучении эффектов аберраций в рамках геометрической оптики (см. гл. 5) мы разлагали функцию аберраций Ф в степенной ряд. Здесь же, поскольку интегрирование производится по единичному кругу, удобнее разлагать Ф по полной системе полиномов, ортогональных внутри единичного круга *). Существует много систем полиномов, обладающих таким свойством; однако одна из них, введенная Цернике [23], обладает еще и простыми свойствами инвариантности. В приложении 7 дан вывод *круговых полиномов* Цернике и обсуждаются некоторые их свойства; здесь мы приведем лишь те формулы, которые потребуются в настоящей главе.

Круговые полиномы Цернике представляют собой полиномы $V_n^l(X, Y)$ от двух действительных переменных X, Y; если выразить X, Y в полярных координатах ($X = \rho\sin\theta$, $Y = \rho\cos\theta$), то полиномы имеют вид

$$V_n^l(\rho\sin\theta,\ \rho\cos\theta) = R_n^l(\rho)\exp(il\theta), \tag{1}$$

где $l \gtrless 0$ и $n \geqslant 0$ — целые числа, $n \geqslant |l|$, а $n - |l|$ — четное число. Ортогональность и нормировка полиномов выражаются формулами

$$\iint\limits_{X^2+Y^2\leqslant 1} [V_n^l(X, Y)]^* V_{n'}^{l'}(X, Y)\,dX\,dY = \frac{\pi}{n+1}\delta_{ll'}\delta_{nn'}, \tag{2}$$

где δ_{ij} — символ Кронекера, а звездочка обозначает комплексное сопряжение. Радиальные функции $R_n^l(\rho)$ являются полиномами по ρ, содержащими степени ρ^n, ρ^{n-2}, ..., $\rho^{|l|}$, и, как показано в приложении 7, они тесно связаны с полиномами Якоби (вырожденными гипергеометрическими функциями). Как видно из (1) и (2), радиальные полиномы удовлетворяют соотношениям

$$\int_0^1 R_n^l(\rho) R_{n'}^l(\rho)\,\rho\,d\rho = \frac{1}{2(n+1)}\delta_{nn'}. \tag{3}$$

Они определяются следующими формулами ($m = |l|$):

$$R_n^{\pm m}(\rho) = \frac{1}{\left(\frac{n-m}{2}\right)!\,\rho^m}\left\{\frac{d}{d(\rho^2)}\right\}^{\frac{n-m}{2}}\left\{(\rho^2)^{\frac{n+m}{2}}(\rho^2-1)^{\frac{n-m}{2}}\right\} \tag{4}$$

или

$$R_n^{\pm m}(\rho) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\rho^{n-2s}. \tag{5}$$

Нормировка выбрана так, чтобы при всех возможных значениях n и m

$$R_n^{\pm m}(1) = 1. \tag{6}$$

Радиальные полиномы имеют образующую функцию

$$\frac{\left[1+z-\sqrt{1-2z(1-2\rho^2)+z^2}\right]^m}{(2z\rho)^m\sqrt{1-2z(1-2\rho^2)+z^2}} = \sum_{s=0}^{\infty} z^s R_{m+2s}^{\pm m}(\rho). \tag{7}$$

При $m = 0$ правая часть (7) превращается в образующую функцию полиномов Лежандра **) от аргумента $2\rho^2 - 1$, т. е.

$$R_{2n}^0(\rho) = P_n(2\rho^2 - 1). \tag{8}$$

В табл. 9.1 приведены в явном виде радиальные полиномы для нескольких первых значений индексов.

*) Термин «полная система» означает, что любая, достаточно гладкая функция разлагается в ряд по функциям данной системы. Более точное определение этого термина приводится, например, в книге [22].

**) См., например, [22].

Таблица 9.1

Радиальные полиномы $R_n^m(\rho)$ для $m \leqslant 8$, $n \leqslant 8$

m \ n	0	1	2	3	4	5	6	7	8
0	1		$2\rho^2-1$		$6\rho^4-6\rho^2+1$		$20\rho^6-30\rho^4+12\rho^2-1$		$70\rho^8-140\rho^6+90\rho^4-20\rho^2+1$
1		ρ		$3\rho^3-2\rho$		$10\rho^5-12\rho^3+3\rho$		$35\rho^7-60\rho^5+30\rho^3-4\rho$	
2			ρ^2		$4\rho^4-3\rho^2$		$15\rho^6-20\rho^4+6\rho^2$		$56\rho^8-105\rho^6+60\rho^4-10\rho^2$
3				ρ^3		$5\rho^5-4\rho^3$		$21\rho^7-30\rho^5+10\rho^3$	
4					ρ^4		$6\rho^6-5\rho^4$		$28\rho^8-42\rho^6+15\rho^4$
5						ρ^5		$7\rho^7-6\rho^5$	
6							ρ^6		$8\rho^8-7\rho^6$
7								ρ^7	
8									ρ^8

В теории Нижбера — Цернике важную роль играет следующее соотношение (также доказанное в приложении):

$$\int_0^1 R_n^m(\rho)\, J_m(v\rho)\, \rho\, d\rho = (-1)^{\frac{n-m}{2}} \frac{J_{n+1}(v)}{v}, \qquad (9)$$

где J — функция Бесселя первого рода.

Вместо комплексных полиномов V можно использовать вещественные полиномы

$$\left.\begin{aligned} U_n^m &= \frac{1}{2}\left[V_n^m + V_n^{-m}\right] = R_n^m(\rho)\cos m\theta, \\ U_n^{-m} &= \frac{1}{2i}\left[V_n^m - V_n^{-m}\right] = R_n^m(\rho)\sin m\theta. \end{aligned}\right\} \qquad (10)$$

Мы будем в дальнейшем применять только полиномы $U_n^m = R_n^m(\rho)\cos m\theta$, так как волновые искажения симметричны относительно меридиональной плоскости $\theta = 0$, и, следовательно, функция аберраций является четной функцией θ.

9.2.2. Разложение функции аберраций. Следуя Нижберу, разложим функцию аберраций Φ по круговым полиномам Цернике Из симметрии задачи вытекает, как и в § 5.1, что разложение содержит лишь комбинации переменных, а именно Y_1^{*2}, ρ^2 и $Y_1^*\rho\cos\theta$, так что оно должно иметь вид

$$\Phi(Y_1^*, \rho, \theta) = \sum_l \sum_n \sum_m a_{lnm} (Y_1^*)^{2l+m} R_n^m(\rho)\cos m\theta, \qquad (11)$$

где l, n и m — неотрицательные целые числа, $n \geqslant m$, $n-m$ — четное число и a_{lnm} — постоянные.

Поскольку нас будет интересовать главным образом дифракционное изображение фиксированной точки предмета ($Y_1^* = \mathrm{const}$), удобно не выделять явную зависимость Φ от Y_1^* и переписать (11) в виде

$$\Phi = A_{00} + \frac{1}{\sqrt{2}} \sum_{n=2}^{\infty} A_{n0} R_n^0(\rho) + \sum_{n=1}^{\infty} \sum_{m=1}^{n} A_{nm} R_n^m(\rho) \cos m\theta. \tag{12}$$

Коэффициенты A_{nm} зависят от Y_1^*, а множитель $1/\sqrt{2}$ введен перед вторым членом для упрощения окончательных формул.

Если аберрации достаточно малы, то с помощью коэффициентов A можно выразить нормированную интенсивность в параксиальном фокусе в простой форме. Подставляя (12) в (9.1.21) и используя условие ортогональности (3), получим

$$\overline{\Phi} = A_{00}, \qquad \overline{\Phi^2} = A_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} \sum_{m=0}^{n} \frac{A_{nm}^2}{n+1}. \tag{13}$$

Из первого соотношения следует, что A_{00} характеризует среднее запаздывание волнового фронта относительно опорной сферы Гаусса. Второе соотношение представляет собой «формулу Парсеваля» для ортогональной системы функций $R_n^m(\rho)\cos m\theta$. Подставляя (12) в уравнение (9.1.22), получим для нормированной интенсивности в параксиальном фокусе

$$i(P_1^*) = 1 - \frac{2\pi^2}{\lambda^2} \sum_{n=1}^{\infty} \sum_{m=0}^{n} \frac{A_{nm}^2}{n+1}. \tag{14}$$

При решении важной проблемы «сбалансирования» аберраций различных порядков друг относительно друга с целью получения максимальной интенсивности проявляются преимущества разложения по круговым полиномам. Предположим, что аберрация описывается одним членом «степенного ряда», а именно

$$\Phi = A'_{nm} \rho^n \cos^m \theta, \tag{15}$$

где A' — малая постоянная (порядка длины волны или меньше). Спрашивается, можно ли увеличить интенсивность $i(P_1^*)$, если ввести аберрации более низких порядков. Или, более точно, можно ли так подобрать постоянные A'_{pq} в выражении

$$\Phi' = A'_{nm} \rho^n \cos^m \theta + \sum_{p<n} \sum_{q \leqslant p} A'_{pq} \rho^p \cos^q \theta, \tag{16}$$

чтобы интенсивность в параксиальном фокусе была как можно больше.

При любом выборе постоянных A'_{pq} функцию аберрации (16) можно также представить с помощью круговых полиномов

$$\Phi' = \varepsilon_{nm} A_{nm} R_n^m(\rho) \cos m\theta + \sum_{p<n} \sum_{q \leqslant p} \varepsilon_{pq} A'_{pq} R_p^q(\rho) \cos q\theta, \tag{17}$$

где

$$\left.\begin{aligned} \varepsilon_{nm} &= \frac{1}{\sqrt{2}}, \quad \text{когда } m=0,\ n \neq 0, \\ \varepsilon_{nm} &= 1 \qquad \text{в остальных случаях.} \end{aligned}\right\} \tag{18}$$

Коэффициент при наивысшей степени ρ в (16), т. е. при ρ^n, равен, согласно (5), $n!\Big/\left[\frac{1}{2}(n+m)\right]!\left[\frac{1}{2}(n-m)\right]!$; тогда, сравнивая коэффициенты при ρ^n в (16) и (17), получим

$$\frac{n!}{\left[\frac{1}{2}(n+m)\right]!\left[\frac{1}{2}(n-m)\right]!} \varepsilon_{nm} A_{nm} = A'_{nm}. \tag{19}$$

Если коэффициент A'_{nm}, а следовательно, и A_{nm} фиксирован, то, согласно (14), максимальная интенсивность в точке P_1^* получается, когда все коэффициенты под знаком суммы в (17) тождественно равны нулю. Тогда функция аберраций принимает вид

$$\Phi' = \varepsilon_{nm} A_{nm} R_n^m(\rho) \cos m\theta, \tag{20}$$

а интенсивность в P_1^* становится равной

$$i(P_1^*) = 1 - \frac{2\pi^2}{\lambda^2} \frac{A_{nm}^2}{n+1}, \tag{21}$$

где A_{nm} выражается через A'_{nm} с помощью (19). Теперь очевидно, что в *одном аберрационном члене $A_{nm}R_n^m(\rho)\cos m\theta$ разложения* (12) *несколько членов вида* $A'_{pq}\rho^p\cos^q\theta$ с $p = n,\ n-2, \ldots, m$, и $q = m,\ m-2, \ldots 1$ *или* 0 *скомбинировалось таким образом, что при данном (достаточно малом) значении коэффициента при* $\rho^n\cos^m\theta$ *нормированная интенсивность в параксиальном фокусе максимальна.*

Проиллюстрируем полученный результат простым примером. Предположим, что оптическая система создает небольшую сферическую аберрацию шестого порядка $[\Phi = A'_{60}\rho^6]$, но мы можем ввести контролируемые величины сферической аберрации четвертого порядка $[A'_{40}\rho^4]$ и дефокусировки $[A'_{20}\rho^2]$. Требуется найти значения A'_{40} и A'_{20}, при которых интенсивность в дифракционном фокусе максимальна. Задача такого рода была впервые рассмотрена Рихтером [24], который показал, что максимум интенсивности достигается при

$$\frac{A'_{40}}{A'_{60}} = -\frac{3}{2}, \qquad \frac{A'_{20}}{A'_{60}} = \frac{3}{5}. \tag{22}$$

Если взглянуть на табл. 9.1, то мы увидим, что эти отношения в точности равны отношениям соответствующих коэффициентов полинома $R_6^0(\rho)$, т. е.

$$R_6^0(\rho) = 20\rho^6 - 30\rho^4 + 12\rho^2 - 1. \tag{23}$$

Таким образом, в случае достаточно малых аберраций введение круговых полиномов Цернике автоматически решает задачу «сбалансирования» аберраций в указанном смысле; более того, с помощью теоремы смещения можно определить положение дифракционного фокуса.

§ 9.3. Допустимые величины первичных аберраций

Прежде чем исследовать сложную проблему нахождения распределения интенсивности в дифракционном изображении при наличии аберраций, рассмотрим более простую задачу оценки максимальной величины аберраций, допустимой в оптической системе.

Из предыдущего обсуждения следует, что при наличии аберраций максимальная интенсивность в дифракционном изображении меньше интенсивности в параксиальном фокусе (центре картины Эйри) оптической системы с теми же апертурой и фокусным расстоянием, но свободной от аберраций. Рэлей [1] впервые показал, что интенсивность света в параксиальном фокусе падает меньше чем на 20% (такая потеря обычно допустима), если первичная сферическая аберрация в системе такова, что волновой фронт в выходном зрачке отстоит от опорной сферы Гаусса на расстоянии, меньшем четверти длины волны. Более поздние исследователи установили, что качество изображения при наличии других обычно встречающихся аберраций существенно не ухудшается, если деформация волнового фронта не превышает четверти длины волны. Полученный результат известен как *правило четверти волны Рэлея*, служащее полезным критерием допустимой величины аберраций в оптической системе, формирующей изображение. Это правило служит, конечно, лишь грубым указанием на необходимость коррекции системы, поскольку распределение света в изображе-

нии зависит не только от максимальной деформации волновых фронтов, но и от их формы (типа аберрации). Более того, допустимое количество потерь света зависит также от назначения оптических инструментов, и поэтому в некоторых случаях приходится вводить более строгие допуски.

Если условие $|\Phi_{\text{макс}}|=\lambda/4$ применить к аберрациям разного типа, то получающиеся значения интенсивности в дифракционном фокусе окажутся различными. Поэтому удобнее сформулировать такой критерий, который соответствует заданному значению интенсивности в дифракционном фокусе. Критерий подобного типа рассмотрел Марешаль [25], воспользовавшись соотношением между интенсивностью в центре опорной сферы и среднеквадратичным отклонением волнового фронта от сферической формы.

Если аберрации достаточно малы, то, согласно (9.1.24), для интенсивности в точке P в области изображения имеем

$$i(P) \approx 1-\left(\frac{2\pi}{\lambda}\right)^2 (\Delta\Phi_P)^2. \tag{1}$$

Следуя Марешалю, *мы будем считать, что система хорошо скорректирована, если нормированная интенсивность в дифракционном фокусе F больше или равна* 0,8. Из (1) следует, что $i(F)\geqslant 0{,}8$, когда $|\Delta\Phi_F|\leqslant\lambda/14$, т. е. приведенное выше условие эквивалентно требованию, чтобы *среднеквадратичное отклонение волнового фронта от опорной сферы, центр которой находится в дифракционном фокусе, не превышало* $\lambda/14$.

Фактически условие Марешаля вытекает из несколько иного неравенства. Из (9.1.20) следует, что

$$i(F)\geqslant\frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi}\exp(ik\Phi_F)\,\rho\,d\rho\,d\theta\right|^2\geqslant\frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi}\cos(k\Phi_F)\,\rho\,d\rho\,d\theta\right|^2.$$

Если $k|\Phi_F|<\pi/2$, т. е. $|\Phi_F|<\lambda/4$, то в этом неравенстве величину $\cos(k\Phi_F)$ можно заменить на $1-{}^1/_2(k\Phi_F)^2$. Если, кроме того, выбрать радиус опорной сферы так, чтобы $\overline{\Phi}_F=0$, то $(\Delta\Phi_F)^2=\overline{\Phi_F^2}$, и неравенство примет вид

$$i(P)\geqslant\left[1-\frac{2\pi^2}{\lambda^2}(\Delta\Phi_F)^2\right]^2. \tag{1a}$$

Критерий Марешаля следует из неравенства (1а), однако если аберрации малы, то он практически не отличается от приведенного выше критерия. Для наших целей удобнее пользоваться соотношением (1), а не (1а), поскольку первый более прямым образом связан с экстремальными свойствами круговых полиномов Цернике.

Определим теперь положение дифракционного фокуса и допуски для первичных аберраций (аберраций Зайделя). В обозначениях, принятых в настоящей главе, каждая первичная аберрация представляет собой деформацию волнового фронта вида *)

$$\Phi = a'_{lnm}(Y_1^*)^{2l+m}\rho^n\cos^m\theta, \tag{2}$$

где $2l+m+n=4$. Удобно положить

$$A'_{lnm}=a'_{lnm}(Y_1^*)^{2l+m}, \tag{3}$$

и тогда (2) принимает вид

$$\Phi = A'_{lmn}\rho^n\cos^m\theta. \tag{4}$$

Постоянные A' легко выразить через коэффициенты Зайделя B, C, D, E и F, введенные в гл. 5. Если произвольную постоянную λ_0, входящую в (5.2.7) и (5.2.8), положить равной единице, то λ_1 обозначает увеличение плоскость входного — плоскость выходного зрачков; далее, если вспомнить, что коэффициент n_1 равен теперь единице, то переменным ρ и y_0 из (5.3.7) соответствуют

*) Штрихами отмечаются коэффициенты степенных рядов, а нештрихованные коэффициенты относятся к разложению по полиномам Цернике.

величины $a\rho/\lambda_1$ и $-\lambda_1 Y_1^*/R$ из настоящего раздела. Сравнивая (4) и (5.3.7), получим

$$\left.\begin{aligned} A'_{040} &= -\frac{1}{4}\left(\frac{a}{\lambda_1}\right)^4 B, \quad A'_{031} = -\left(\frac{a}{\lambda_1}\right)^3\left(\frac{\lambda_1 Y_1^*}{R}\right)F, \\ A'_{022} &= -\left(\frac{a}{\lambda_1}\right)\left(\frac{\lambda_1 Y_1^*}{R}\right)^2 C, \quad A'_{120} = -\frac{1}{2}\left(\frac{a}{\lambda_1}\right)^2\left(\frac{\lambda_1 Y_1^*}{R}\right)^2 D, \\ A'_{111} &= -\left(\frac{a}{\lambda_1}\right)\left(\frac{\lambda_1 Y_1^*}{R}\right)^3 E. \end{aligned}\right\} \qquad (5)$$

В выражении через круговые полиномы типичный член, описывающий аберрацию, имеет вид *)

$$\Phi = \varepsilon_{nm} A_{lnm} R_n^m(\rho)\cos m\theta. \qquad (6)$$

В последней колонке табл. 9.2 выписаны члены с индексами l, m и n, удовлетворяющими равенству $2l+m+n=4$ (первичные аберрации) Как мы видим, некоторые члены Зайделя сопровождаются теперь членами более низкого порядка, что в соответствии с теоремой смещения (см. п. 9.1.2) вызывает сдвиг распределения интенсивности как целого

Согласно теореме, приведенной в п 9 2 2, максимальная интенсивность в изображении, искаженном аберрацией типа (6), достигается в параксиальном фокусе. Следовательно, сравнивая соответствующие члены в двух последних колонках табл. 9.2, можно сразу же определить координаты дифракционного

Таблица 9.2

Представление первичных аберраций

Тип аберрации	l	n	m	Представление в форме (4)	Представление в форме (6)
Сферическая аберрация	0	4	0	$A'_{040}\rho^4$	$\frac{1}{\sqrt{2}}A'_{040}R_4^0(\rho)=\frac{1}{\sqrt{2}}A_{040}(6\rho^4-6\rho^2+1)$
Кома	0	3	1	$A'_{031}\rho^3\cos\theta$	$A_{031}R_3^1(\rho)\cos\theta = A_{031}(3\rho^3-2\rho)\cos\theta$
Астигматизм	0	2	2	$A'_{022}\rho^2\cos^2\theta$	$A_{022}R_2^2(\rho)\cos 2\theta = A_{022}\rho^2(2\cos^2\theta-1)$
Кривизна поля	1	2	0	$A'_{120}\rho^2$	$\frac{1}{\sqrt{2}}A_{120}R_2^0(\rho)=\frac{1}{\sqrt{2}}A_{120}(2\rho^2-1)$
Дисторсия	1	1	1	$A'_{111}\rho\cos\theta$	$A_{111}R_1^1(\rho)\cos\theta = A_{111}\rho\cos\theta$

фокуса изображения, искаженного первичной аберрацией. Проиллюстрируем последнее утверждение, подробно рассмотрев случай сферической аберрации. Эта аберрация описывается членом

$$\Phi = A'_{040}\rho^4. \qquad (7)$$

*) Для обеспечения единообразия с формулами, приведенными в § 9 2, здесь оставлен множитель ε_{nm}, равный $1/\sqrt{2}$ при $m=0$, $n\neq 0$ и 1 при остальных значениях m и n.

Соответствующее выражение в форме (6) имеет вид

$$\Phi=\frac{1}{\sqrt{2}}A_{040}R_4^0(\rho)=\frac{1}{\sqrt{2}}A_{040}(6\rho^4-6\rho^2+1). \tag{8}$$

Если положить

$$A'_{040}=\frac{6}{\sqrt{2}}A_{040}, \tag{9}$$

то распределения интенсивности в обоих случаях будут, согласно теореме смещения, одинаковы; однако распределение, соответствующее (7), будет сдвинуто относительно распределения, соответствующего (8), на расстояние, определяемое из уравнения (9.1.15) с

$$H=\frac{6A_{040}}{\sqrt{2}}=A'_{040},\qquad K=L=0,\qquad M=-\frac{A_{040}}{\sqrt{2}}=\frac{A'_{040}}{6},$$

т. е.

$$x'=x,\qquad y'=y,\qquad z'=z+2(R/a)^2A'_{040}. \tag{10}$$

Так как максимум интенсивности в дифракционном изображении, соответствующем (8), находится в начале координат $x=y=z=0$, то дифракционный фокус F при наличии первичной сферической аберрации типа (7) расположен в точке

$$x_F=y_F=0,\qquad z_F=2(R/a)^2A'_{040}. \tag{11}$$

Координаты точки F, определяемые (11), допускают простое геометрическое истолкование Пусть ΔY и ΔZ — поперечная и продольная сферические аберрации, считающиеся положительными, если луч пересекает ось с положительной стороны от параксиального фокуса. Из (5.1.16) при $\Phi=A'_{040}\rho^4$, $\rho=Y/a$, $D_1\approx-R\approx-R'$, $X_0=X=0$, $n_1=1$ имеем

$$\Delta Y=Y_1-Y_1^*=4\left(\frac{R}{a}\right)\left(\frac{Y}{a}\right)^3A'_{040}, \tag{12a}$$

следовательно, на основании элементарных геометрических соображений и предыдущего соотношения

$$\Delta Z=Z_1-Z_1^*\approx\frac{R}{Y}\Delta Y=4\left(\frac{R}{a}\right)^2\left(\frac{Y}{a}\right)^2A'_{040}. \tag{12б}$$

Для краевого луча ($Y=a$) это дает $(\Delta Z)_{макс}=4(R/a)^2A_{040}$. Тогда из (11) следует, что *) *дифракционный фокус при наличии малой первичной сферической аберрации расположен посередине между параксиальным и краевым фокусами.*

Определим теперь максимально допустимую величину сферической аберрации. Согласно (9.2.14) нормированная интенсивность в параксиальном фокусе для любой аберрации типа (6) больше или равна 0,8, если

$$1-\frac{2\pi^2}{\lambda^2}\frac{A_{lnm}^2}{n+1}\geqslant 0{,}8,$$

т. е.

$$|A_{lnm}|\leqslant\frac{\lambda\sqrt{n+1}}{10}. \tag{13}$$

В частности, для первичной сферической аберрации имеем

$$|A_{040}|\leqslant 0{,}22\lambda,$$

или с учетом (9) —

$$|A'_{040}|\leqslant 0{,}94\lambda. \tag{14}$$

Это и есть требуемое условие, определяющее максимально допустимую первичную сферическую аберрацию, согласно которому максимальное отклонение волнового фронта от опорной сферы Гаусса должно быть меньше 0,94 λ.

Совершенно таким же способом можно найти координаты дифракционных фокусов и максимально допустимые первичные аберрации других типов.

*) Здесь мы пренебрегаем теми небольшими эффектами, которые связаны с разным выбором направления z в настоящем разделе и в § 5.1.

В частности, дифракционный фокус при наличии небольшого первичного астигматизма расположен в точке с координатами

$$x_F = y_F = 0, \qquad z_F = \left(\frac{R}{a}\right)^2 A'_{022}. \tag{15}$$

Этот результат также допускает простое физическое истолкование. Согласно (5) и (5 3 18) радиусы тангенциальной и сагиттальной фокальных поверхностей R_t и R_s определяются из выражений (n_1 снова полагается равным единице)

$$\frac{1}{R_t} = -\frac{4}{a^2}(R/Y_1^*)^2 A'_{022}, \qquad \frac{1}{R_s} = 0, \tag{16}$$

так что абсциссы z_t и z_s двух фокальных линий равны

$$z_t = -\frac{(Y_1^*)^2}{2R_t} = 2(R/a)^2 A'_{022}, \qquad z_s = 0. \tag{17}$$

Таким образом, из (15) следует, что *дифракционный фокус при наличии небольшого первичного астигматизма находится посередине между тангенциальной и сагиттальной фокальными линиями.*

Поскольку первичные кривизна поля и дисторсия описываются членами, содержащими соответственно ρ^2 и ρ, их эффект, согласно теореме смещения, состоит лишь в смещении как целого трехмерного распределения интенсивности в свободном от аберраций изображении. Итак, при наличии небольшой первичной кривизны поля или первичной дисторсии нормированная интенсивность i в дифракционном фокусе равна единице, но сам дифракционный фокус не совпадает с параксиальным фокусом.

В табл. 9 3 приведены результаты, относящиеся к первичным аберрациям.

Таблица 9 3

Координаты дифракционных фокусов и условия, определяющие максимально допустимую величину первичных аберраций

Тип аберрации	Координаты дифракционного фокуса F			Условие [$i(F) \geqslant 0{,}8$]
	x_F	y_F	z_F	
Сферическая аберрация	0	0	$2\left(\frac{R}{a}\right)^2 A'_{040}$	$\lvert A'_{040}\rvert \leqslant 0{,}94\lambda$
Кома	0	$\frac{2}{3}\left(\frac{R}{a}\right) A'_{031}$	0	$\lvert A'_{031}\rvert \leqslant 0{,}60\lambda$
Астигматизм	0	0	$\left(\frac{R}{a}\right)^2 A'_{022}$	$\lvert A'_{022}\rvert \leqslant 0{,}35\lambda$
Кривизна поля	0	0	$2\left(\frac{R}{a}\right)^2 A'_{120}$	—
Дисторсия	0	$\left(\frac{R}{a}\right) A'_{111}$	0	—

§ 9.4. Дифракционная картина, получающаяся при наличии одной аберрации

Рассмотрим теперь дифракционное изображение при наличии аберрации, которая описывается одним членом разложения (9.2.11), а именно

$$\Phi = a_{lnm} (Y_1^*)^{2l+m} R_n^m (\rho) \cos m\theta. \quad (1)$$

Как и раньше, избавимся от явной зависимости Φ от Y_1^*, положив

$$\alpha_{lnm} = \frac{2\pi}{\lambda} a_{lnm} (Y_1^*)^{2l+m} = \frac{2\pi}{\lambda} \varepsilon_{nm} A_{lnm}. \quad (2)$$

Введем обозначение

$$C = -\frac{i\pi A}{\lambda} \left(\frac{a}{R}\right)^2 \exp\left[i \left(\frac{R}{a}\right)^2 u\right]. \quad (3)$$

Тогда дифракционный интеграл (9.1.6) принимает вид

$$U(u, v, \psi) = \frac{C}{\pi} \int_0^1 \int_0^{2\pi} \exp\left\{i \left[-v\rho \cos(\theta - \psi) - \frac{1}{2} u\rho^2 + \alpha_{lnm} R_n^m (\rho) \cos m\theta\right]\right\} \rho \, d\rho \, d\theta. \quad (4)$$

Интеграл (4) можно представить в виде бесконечного ряда, если разложить $\exp[-iv\rho \cos(\theta - \psi)]$ и $\exp[i\alpha_{lnm} R_n^m(\rho) \cos m\theta]$ с помощью тождества Якоби (см. (8.8.29))

$$\exp(iz \cos\varphi) = J_0(z) + 2 \sum_{s=1}^{\infty} i^s J_s(z) \cos s\varphi. \quad (5)$$

Перемножая оба ряда, получим

$$\exp\{i[-v\rho \cos(\theta - \psi) + \alpha_{lnm} R_n^m (\rho) \cos m\theta]\} = 4 \sum_{s=0}^{\infty}{}' \sum_{s'=0}^{\infty}{}' (-i)^s (-i)^{s'} J_s [\alpha_{lnm} R_n^m (\rho)] J_{s'} (v\rho) \cos ms\theta \cos[s'(\theta - \psi)], \quad (6)$$

где штрихи при символах суммы означают, что члены с индексами $s=0$ и $s'=0$ следует брать с коэффициентом $1/2$. Подставив (6) в (4) и проинтегрировав почленно по углу θ, находим

$$U(u, v, \psi) = 4C \sum_{s=0}^{\infty}{}' (-i)^{(m-1)s} \cos ms\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] J_s [\alpha_{lmn} R_n^m (\rho)] J_{ms} (v\rho) \rho \, d\rho, \quad (7)$$

где член с индексом $s=0$ снова надо брать с коэффициентом $1/2$.

Поскольку нас интересуют только малые аберрации (α мало), можно разложить член $J_s[\alpha_{lnm} R_n^m(\rho)]$ под знаком интеграла в степенной ряд, а затем расположить его члены по возрастающим степеням α_{lnm}. Тогда имеем

$$U(u, v, \psi) = C[U_0 + i\alpha_{lnm} U_1 + (i\alpha_{lnm})^2 U_2 + (i\alpha_{lnm})^3 U_3 + (i\alpha_{lnm})^4 U_4 + \ldots], \quad (8a)$$

где

$$\left.\begin{aligned}
U_0 &= 2\int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] J_0(v\rho)\,\rho\,d\rho,\\
U_1 &= 2(-i)^m \cos m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] R_n^m(\rho)\, J_m(v\rho)\,\rho\,d\rho,\\
U_2 &= \frac{1}{2!}\left\{\int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^2 J_0(v\rho)\,\rho\,d\rho +\right.\\
&\qquad \left. + i^{2m}\cos 2m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^2 J_{2m}(v\rho)\,\rho\,d\rho\right\},\\
U_3 &= \frac{1}{2\cdot 3!}\left\{3(-i)^m \cos m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^3 J_m(v\rho)\,\rho\,d\rho +\right.\\
&\qquad \left. + (-i)^{3m}\cos 3m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^3 J_{3m}(v\rho)\,\rho\,d\rho\right\},\\
U_4 &= \frac{1}{2^2 4!}\left\{3\int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^4 J_0(v\rho)\,\rho\,d\rho +\right.\\
&\qquad + 4i^{2m}\cos 2m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^4 J_{2m}(v\rho)\,\rho\,d\rho +\\
&\qquad \left. + i^{4m}\cos 4m\psi \int_0^1 \exp\left[-\frac{1}{2} iu\rho^2\right] \{R_n^m(\rho)\}^4 J_{4m}(v\rho)\,\rho\,d\rho\right\}.
\end{aligned}\right\} \tag{8б}$$

Как показал Нижбер, если u и α_{lnm} равны по порядку величины единице, то первых четырех членов разложения (8а) достаточно для получения интенсивности с точностью до нескольких процентов.

Для вычисления интегралов в (8б) можно поступить следующим образом. Выразим множитель $\exp\left[-{}^1/_2\, iu\rho^2\right]$ через радиальные полиномы с помощью хорошо известной формулы Бауэра *) в виде

$$\exp(iz\cos\varphi) = \left(\frac{\pi}{2z}\right)^{1/2} \sum_{s=0}^{\infty} i^s (2s+1)\, J_{s+1/2}(z)\, P_s(\cos\varphi), \tag{9}$$

где P_s — полиномы Лежандра. Полагая $\cos\varphi = 2\rho^2 - 1$ и используя соотношение $P_s(2\rho^2-1) = R_{2s}^0(\rho)$ (см. (9.2.8)), получим

$$\exp(-{}^1/_2\, iu\rho^2) = \exp(-{}^1/_4\, iu)\exp(-{}^1/_4\, iu(2\rho^2-1)) =$$
$$= \exp(-{}^1/_4\, iu)\sqrt{\frac{2\pi}{u}} \sum_{s=0}^{\infty} (-i)^s (2s+1)\, J_{s+1/2}({}^1/_4\, u)\, R_{2s}^0(\rho). \tag{10}$$

Если подставить (10) в (8б), то в правые их части будут входить интегралы, каждый из которых состоит из функции Бесселя, умноженной на произведение радиальных полиномов. Эти интегралы можно вычислить с помощью формулы (9.2.9), а именно

$$\int_0^1 R_n^m(\rho)\, J_m(v\rho)\,\rho\,d\rho = (-1)^{\frac{n-m}{2}} \frac{J_{n+1}(v)}{v}, \tag{11}$$

*) См., например, [26].

при условии, что каждое произведение радиальных полиномов выражается в виде линейной комбинации членов вида $\sum_p A_p R_p^m(\rho)$ с индексом m, равным порядку функции Бесселя, на которую умножается данное произведение. Общее выражение для коэффициентов A_p получить довольно трудно, но если m и n не очень велики, то требуемые линейные комбинации строятся довольно просто с помощью табл. 9.1, что будет показано ниже на нескольких примерах. Для ознакомления с методами, применяемыми в более общих случаях, мы отсылаем читателя к работе Нижбера.

В частном случае волны, свободной от аберраций, подстановка (10) в (7) дает с учетом (11)

$$U(u,\ v,\ \psi)=2C\exp\left(-{}^1/_4\, iu\right)\sqrt{\frac{2\pi}{u}}\sum_{s=0}(-i)^s(2s+1)\,J_{s+1/2}\left({}^1/_4\, u\right)\times$$

$$\times\int_0^1 R_{2s}^0(\rho)\,J_0(v\rho)\,\rho\,d\rho=$$

$$=2C\exp\left(-{}^1/_4\, iu\right)\sqrt{\frac{2\pi}{u}}\sum_{s=0}^{\infty}(i)^s(2s+1)\,J_{s+1/2}\left({}^1/_4\, u\right)\frac{J_{2s+1}(v)}{v}. \qquad (12)$$

Полученное разложение, несмотря на формальное различие, эквивалентно рядам Ломмеля, описанным в п. 8.8.1.

Из соотношений (8) можно сразу же получить некоторые общие свойства дифракционного изображения. Как мы видим, величина U не изменяется при замене ψ на $\psi+2\pi\mu/m$ ($\mu=1,\ 2,\ \ldots,\ m$); следовательно, *ось z является осью симметрии m-го порядка*. Более того, *плоскости, проходящие через ось z и составляющие с плоскостью $x=0$ углы $\pi\mu/m$, служат плоскостями симметрии.* При $m=0$ система обладает, конечно, вращательной симметрией.

Рассмотрим теперь симметрию относительно плоскости $z=0$. Отметим, что при замене u на $-u$ все интегралы в (8) превращаются в комплексно сопряженные. Если m — нечетное число, то все коэффициенты при интегралах вещественны. Поэтому $U(u,\ v,\ \psi)$ в этом случае заменяется на комплексно сопряженную величину и, следовательно, интенсивность не изменяется. Иными словами, *распределение интенсивности симметрично относительно плоскости $z=0$, если m — нечетное число.* Если же m — четное, но не равное нулю число, то коэффициенты, содержащие множитель $\cos 2\mu m\psi$ (здесь μ — целое число), оказываются вещественными, а коэффициенты, содержащие множитель $\cos(2\mu+1)m\psi$,— чисто мнимыми. Поэтому величина U превращается в комплексно сопряженную при одновременной замене u на $-u$ и ψ на $\psi+\pi/m$. Следовательно, *если m — четное, но отличное от нуля число, то интенсивность в любой точке плоскости $z=\mathrm{const}$ равна интенсивности в точке, получающейся при зеркальном отражении исходной точки относительно плоскости $z=0$ с последующим поворотом на угол π/m вокруг оси z.* Таким образом, ось z, которая в общем случае, как отмечалось выше, служит осью симметрии m-го порядка, *является при четном m осью симметрии $2m$-го порядка относительно дифракционной картины в плоскости $z=0$.* При $m=0$ (сферическая аберрация) дифракционное изображение несимметрично относительно плоскости $z=0$.

Наконец отметим, что *распределение интенсивности сохраняется при изменении знака коэффициента аберрации α_{lnm} и одновременной замене u на $-u$, когда m четно, или при замене ψ на $\psi+\pi$, когда m нечетно.*

Рассмотрим теперь кратко структуру изображения, получающегося в системе при наличии небольшой первичной аберрации.

9.4.1. Первичная сферическая аберрация. В этом случае $l=n=0$ и $m=4$. Функция аберраций не зависит от θ, и трехмерное дифракционное изображение

обладает вращательной симметрией относительно главного направления $v=0$. Согласно (8) разложение дифракционного интеграла по степеням α имеет вид

$$U(u, v, \psi) = C\,[U_0 + i(\alpha_{040})\,U_1 + (i\alpha_{040})^2\,U_2 + \ldots], \tag{13}$$

где U_0 характеризует возмущение свободного от аберрации изображения, а U_1, U_2, ... определяются другими соотношениями (8б) при $m = n = 0$. В частности,

$$U_1 = 2\int\limits_0^1 \exp(-{}^1/_2\, iu\rho^2)\, R_4^0(\rho)\, J_0(v\rho)\,\rho\, d\rho. \tag{14}$$

Подставляя вместо $\exp(-{}^1/_2\, iu\rho^2)$ выражение (10), получим

$$U_1 = 2\exp(-{}^1/_4\, iu)\sqrt{\frac{2\pi}{u}}\sum_{s=0}^{\infty}(-i)^s(2s+1)\,J_{s+1/2}({}^1/_4\, u)\times$$
$$\times\int\limits_0^1 R_{2s}^0(\rho)\, R_4^0(\rho)\, J_0(v\rho)\,\rho\, d\rho. \tag{15}$$

Чтобы вычислить интеграл, стоящий в правой части, заменим, как уже объяснялось раньше, произведение радиальных полиномов их линейной комбинацией, причем их верхние индексы равны порядку функции Бесселя (в данном случае нулю). Линейную комбинацию легко получить, если воспользоваться выражением (9.2.8)

$$R_{2s}^0(\rho) = P_s(2\rho^2 - 1) \tag{16}$$

и некоторыми хорошо известными соотношениями для полиномов Лежандра. Имеем

$$R_{2s}^0(\rho)\, R_4^0(\rho) = P_s(2\rho^2-1)\, P_2(2\rho^2-1). \tag{17}$$

Далее $P_2(t) = \frac{1}{2}(3t^2 - 1)$; следовательно, правую часть (17) можно представить в виде

$$P_s(t)\, P_2(t) = \frac{3}{2}\, t^2 P_s(t) - \frac{1}{2}\, P_s(t). \tag{18}$$

Применяя дважды рекуррентную формулу *)

$$tP_s(t) = \frac{1}{2s+1}\,[(s+1)\, P_{s+1}(t) + sP_{s-1}(t)], \tag{19}$$

получим

$$P_2(t)\, P_s(t) = a_s P_{s+2}(t) + b_s P_s(t) + c_s P_{s-2}(t), \tag{20}$$

где

$$a_s = \frac{3(s+2)(s+1)}{2(2s+3)(2s+1)}, \quad b_s = \frac{(s+1)s}{(2s+3)(2s-1)}, \quad c_s = \frac{3}{2}\,\frac{s(s-1)}{(2s+1)(2s-1)}. \tag{21}$$

Следовательно, (17) принимает вид

$$R_{2s}^0(\rho)\, R_4^0(\rho) = a_s R_{2s+4}^0(\rho) + b_s R_{2s}^0(\rho) + c_s R_{2s-4}^0(\rho). \tag{22}$$

Подставляя последнее соотношение в (15) и используя (11), окончательно находим

$$U_1 = 2\exp(-{}^1/_4\, iu)\sqrt{\frac{2\pi}{u}}\,\frac{1}{v}\sum_{s=0}^{\infty}(i)^s(2s+1)\,J_{s+1/2}({}^1/_4\, u)\times$$
$$\times\,[a_s J_{2s+5}(v) + b_s J_{2s+1}(v) + c_s J_{2s-3}(v)]. \tag{23}$$

*) См., например, [27].

Таким же образом *) можно разложить в ряд U_2, U_3, ... Используя полученные ряды, можно вычислить значения интенсивности $I = |U|^2$ в ряде точек изображения и построить затем изофоты (линии равной интенсивности).

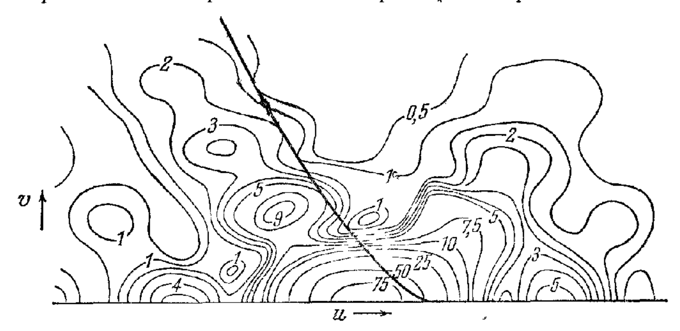


Рис 9 3. Изофоты в меридиональной плоскости при наличии первичной сферической аберрации $\Phi = 0{,}48\lambda\rho^4$ (из [13])

Жирная линия обозначает геометрическую каустику Интенсивность нормирована на 100 в центре изображения, свободного от аберраций Интенсивность Штреля в плоскости наилучшего изображения равна 0,95

В любой плоскости, перпендикулярной к главному направлению $v=0$, изофоты, конечно, имеют вид окружностей.

На рис. 9.3 изображены изофоты в меридиональной плоскости при наличии первичной сферической аберрации $\Phi = 0{,}48\lambda\rho^4$, а на рис. 9.4 и 9 5 воспроизведены фотографии изображений в различных плоскостях при наличии несколько большей сферической аберрации **).

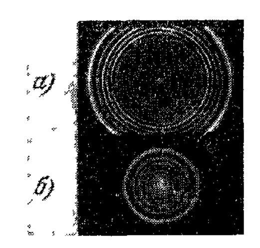


Рис. 9 4. Изображения в плоскости краевого фокуса (*а*) и в плоскости геометрического кружка наименьшего рассеяния (*б*) при наличии первичной сферической аберрации $\Phi = 16\lambda\rho^4$ [15]

9.4.2. Первичная кома. В этом случае $l=0$, $n=3$ и $m=1$. Согласно табл. 9.3 дифракционный фокус находится в плоскости $z=0$, а возмущение в этой плоскости описывается выражением

$$U(0, v, \psi) = C\,[U_0(0, v, \psi) + (i\alpha_{031})\,U_1(0, v, \psi) + (i\alpha_{031})^2\,U_2(0, v, \psi) + \ldots]. \quad (24)$$

При $u=0$ интеграл U_0, определенный (86), характеризует возмущение $2J_1(v)/v$ в фокальной плоскости свободной от аберрации системы (картина Эйри), а U_1 легко вычислить с помощью (11). Для нахождения U_2, U_3, ... мы должны снова выразить произведения круговых полиномов через соответствующую их линейную комбинацию. В частности, с помощью табл. 9.1 можно доказать, что

$$(R_3^1)^2 = \frac{1}{4}R_0^0 + \frac{1}{20}R_2^0 + \frac{1}{4}R_4^0 + \frac{9}{20}R_6^0 = \frac{2}{5}R_2^2 + \frac{3}{5}R_6^2. \quad (25)$$

Используя эти соотношения в выражении для U_2 (86) и применяя (11), легко

*) Подробности см. в [13]

**) Разложение (13) непригодно для вычисления интенсивности, если аберрации не малы по сравнению с длиной волны. При наличии первичной сферической аберрации, равной нескольким длинам волн, изофоты в меридиональной плоскости были вычислены с помощью механического интегратора Марешалем; его результаты опубликованы в [20, 28, 29] (см. также [5]).

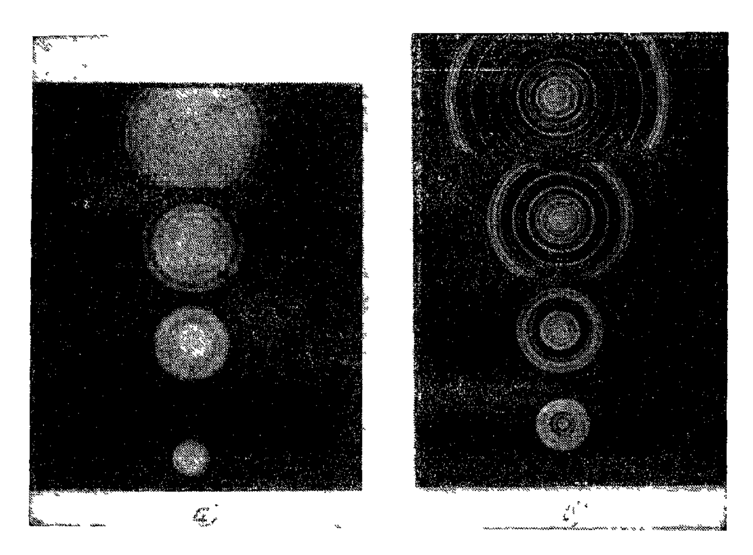


Рис. 9.5. Изображения в параксиальной фокальной плоскости (*а*) и в плоскости геометрического кружка наименьшего рассеяния (*б*) при наличии первичной сферической аберрации $\Phi=17{,}5\lambda\rho^4$; $8{,}4\lambda\rho^4$, $3{,}7\lambda\rho^4$ и $1{,}4\lambda\rho^4$ [15]

Масштаб в *б* втрое больше масштаба в *а*

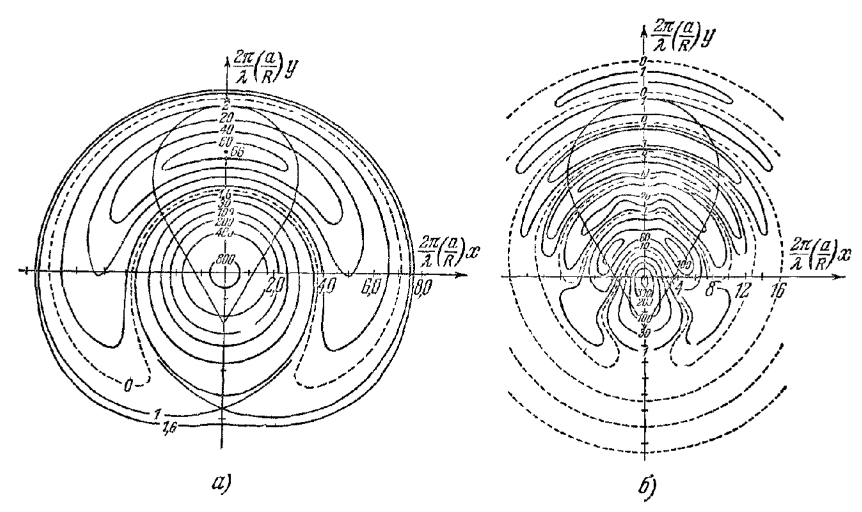


Рис. 9.6. Изофоты в плоскости $z=0$ при наличии первичной комы.

Пунктиром показаны линии нулевой интенсивности. Показана также граница изображения в приближении геометрической оптики. Интенсивность нормирована на 1000 в центре изображения, свободного от аберраций. *а*) $\Phi=0{,}48\lambda\left(\rho^3-\frac{2}{3}\rho\right)\cos\theta$, интенсивность Штреля равна 0,879 [10]; *б*) $\Phi=1{,}4\lambda\left(\rho^3-\frac{2}{3}\rho\right)\cos\theta$, интенсивность Штреля равна 0,306 [14].

вычислить интегралы, входящие в U_2, в результате чего окончательно получим

$$\left.\begin{aligned} &U_0(0, v, \psi) = \frac{2J_1(v)}{v}, \qquad U_1(0, v, \psi) = i\cos\psi\,\frac{2J_4(v)}{v}, \\ &U_2(0, v, \psi) = \frac{1}{2v}\left\{\frac{1}{4}J_1(v) - \frac{1}{20}J_3(v) + \right. \\ &\qquad \left. + \frac{1}{4}J_5(v) - \frac{9}{20}J_7(v) - \cos 2\psi\left(\frac{2}{5}J_3(v) + \frac{3}{5}J_7(v)\right)\right\}. \end{aligned}\right\} \tag{26}$$

На рис. 9.6 и 9.7 показаны изофоты, соответствующие разным величинам первичной комы. Данные на рис. 9.6 получены с помощью разложения в ряды,

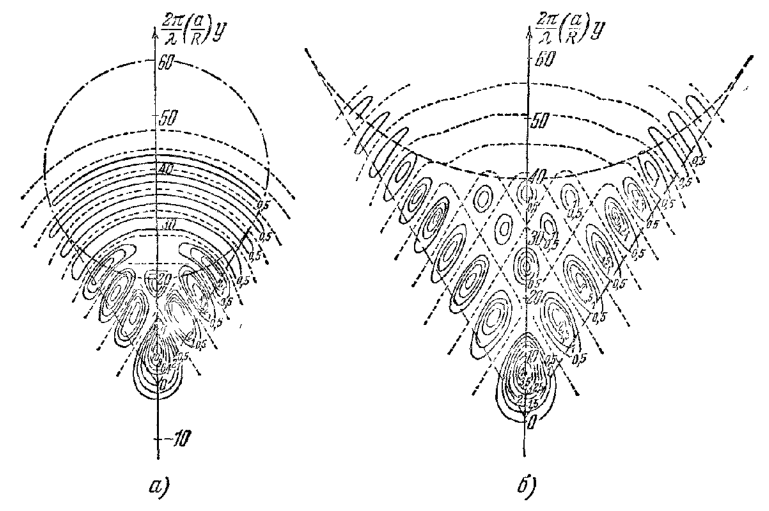


Рис. 9.7. Изофоты в плоскости $z=0$ при наличии первичной комы [30].
Интенсивность нормирована на 100 в центре изображения, свободного от аберраций. *а*) $\Phi = 3.2\ \lambda\rho^3 \cos\theta$; *б*) $\Phi = 6.4\ \lambda\rho^3 \cos\theta$



Рис. 9.8. Изображения в параксиальной фокальной плоскости при наличии комы [15].
$\Phi = 0,3\lambda\rho^3 \cos\theta$, $\lambda\rho^3 \cos\theta$; $2,4\lambda\rho^3 \cos\theta$; $5\lambda\rho^3 \cos\theta$ и $10\rho^3\lambda \cos\theta$

на рис. 9.7 — с помощью численного интегрирования. На рис. 9.8 помещены фотографии изображений, искаженных первичной комой. Приведенные рисунки показывают, что при аберрации порядка длины волны изображение

непохоже ни на картину Эйри, ни на изображение, предсказываемое геометрической оптикой. При увеличении аберрации истинное изображение быстро принимает вид, получающийся в геометрической оптике, однако его пересекает ряд темных полос; можно показать, что они возникают из-за интерференции лучей, дифрагировавших на концах диаметра отверстия.

Рис. 9.6 иллюстрирует также общий результат, установленный в § 9.2 и состоящий в следующем: если малая аберрация описывается членом, выраженным через круговой полином, то распределение интенсивности смещается так, что ее максимум попадает в начало координат.

9.4.3. **Первичный астигматизм.** Аналогичным образом можно исследовать эффект малого первичного астигматизма. В этом случае $l=0$, $n=m=2$, и,

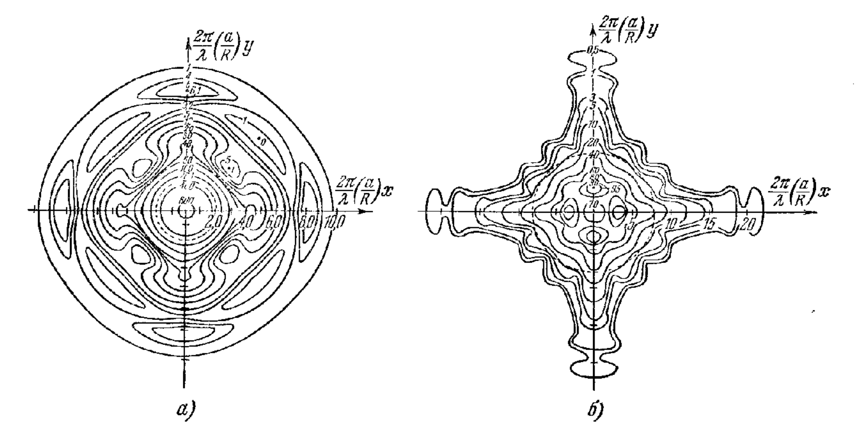


Рис. 9.9. Изофоты в центральной плоскости при наличии первичного астигматизма.

Пунктирные окружности показывают границу изображения в приближении геометрической оптики. Интенсивность нормирована на 1000 в центре изображения, свободного от аберрации. *а*) $\Phi=0{,}16\lambda\rho^2\cos 2\theta$. Интенсивность Штреля равна 0,84 [10]; *б*) $\Phi=0{,}64\lambda\rho^2\cos 2\theta$; интенсивность Штреля равна 0,066 [15].

как было показано в § 9.3, дифракционный фокус находится посередине между двумя фокальными линиями. Рассмотрим распределение света в *центральной плоскости*, т. е. в плоскости, проходящей через указанную точку перпендикулярно к главному направлению. Если аберрация описывается с помощью соответствующего кругового полинома $[A_{022}R_2^2(\rho)\cos 2\theta]$, то центральной плоскостью служит плоскость $u=0$.

Возмущение в центральной плоскости описывается выражением

$$U(0,\ v,\ \psi)=C\,[U_0(0,\ v,\ \psi)+(i\alpha_{022})\,U_1(0,\ v,\ \psi)+(i\alpha_{022})^2\,U_2(0,\ v,\ \psi)+\ldots], \quad (27)$$

где $U_0(0,\ v,\ \psi)$ определяет, как и раньше, распределение интенсивности в картине Эйри, а $U_1(0,\ v,\ \psi)$ легко вычисляется с помощью (11). Для нахождения U_2 используем тождества

$$(R_2^0)^2=\frac{1}{3}R_0^0+\frac{1}{2}R_2^0+\frac{1}{6}R_4^0, \qquad (R_2^2)^2=R_4^4. \quad (28)$$

В результате получим

$$\left.\begin{aligned} U_0(0,\ v,\ \psi)&=\frac{2J_1(v)}{v}, \quad U_1(0,\ v,\ \psi)=-2\cos 2\psi\,\frac{2J_3(v)}{v},\\ U_2(0,\ v,\ \psi)&=\frac{1}{2v}\left\{\frac{1}{3}J_1(v)-\frac{1}{2}J_3(v)+\frac{1}{6}J_5(v)+\cos 4\psi J_5(v)\right\}.\end{aligned}\right\} \quad (29)$$

На рис. 9.9 показаны диаграммы изофот изображений, искаженных астигматизмом. Данные, приведенные на рис. 9.9, *а*, были рассчитаны с помощью

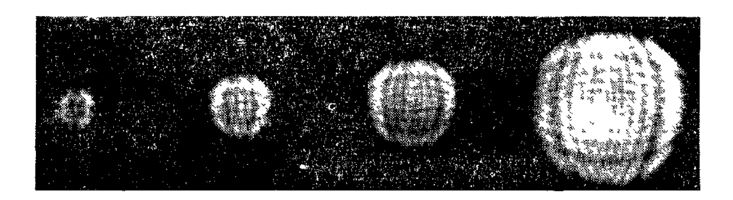


Рис. 9.10. Изображения в центральной плоскости при наличии первичного астигматизма [15]. $\Phi = 1{,}4\lambda\rho^2 \cos 2\theta$, $2{,}7\lambda\rho^2 \cos 2\theta$, $3{,}5\lambda\rho^2 \cos 2\theta$, $6{,}5\lambda\rho^2 \cos 2\theta$.

разложения (27), содержавшего члены вплоть до четвертого порядка по α. На рис. 9.10 и 9.11 показаны фотографии искаженных астигматизмом изображений. Как мы видим, при наличии небольшого астигматизма изофоты в центральной плоскости имеют вид окружностей лишь вблизи центра и более сложную форму на краях изображения. При увеличении астигматизма изображение становится подушкообразным и пересечено интерференционными полосами.

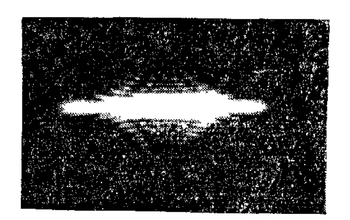


Рис. 9.11. Изображение в плоскости, в которой лежит фокальная линия, при наличии первичного астигматизма $\Phi = 2{,}7\lambda\rho^2 \cos 2\theta$ [15].

Что касается двух оставшихся типов первичной аберрации (кривизны поля и дисторсии), то мы уже показали, что они не нарушают структуру трехмерного изображения, а только смещают положение дифракционного фокуса. Поэтому диаграммы изофот вблизи фокуса при наличии таких аберраций не отличаются от соответствующих диаграмм для свободных от аберраций изображений (см. рис. 8.39), а лишь смещаются относительно параксиального фокуса на величины, указанные в табл. 9.3.

§ 9.5. Изображение протяженных предметов

До сих пор мы изучали только изображения точечных источников. Опишем теперь некоторые общие методы, основанные на фурье-преобразованиях, применяемых при исследовании изображений протяженных предметов. Эти методы были развиты, главным образом в работах Дюффье [31], частично в сотрудничестве с Лансро [32]; в дальнейшем они были развиты и применены к решению частных задач многими исследователями (см., например, [33—37]).

Мы рассмотрим отображения когерентным и некогерентным светом раздельно.

9.5.1. Когерентное освещение. Обозначим декартовы координаты точек в плоскостях параксиального изображения и выходного зрачка соответственно через (x_1, y_1) и (ξ, η) и будем считать, что оси выбранных систем координат взаимно параллельны, а их начала расположены на оси. Точки, лежащие в плоскости предмета, удобно характеризовать такими нормализованными координатами (x_0, y_0), что

$$x_0 = MX_0, \qquad y_0 = MY_0, \tag{1}$$

где (X_0, Y_0) — декартовы координаты произвольной точки предмета, а M — поперечное увеличение; в результате численные значения координат точки предмета и ее параксиального изображения будут одинаковыми *).

Оптическую систему можно охарактеризовать с помощью *функции пропускания* $K(x_0, y_0; x_1, y_1)$, которая определяется как комплексная амплитуда возмущения (рассчитанного на единицу площади в плоскости $(x_0 y_0)$) в точке (x_1, y_1) в плоскости параксиального изображения, обусловленного возмущением с единичной амплитудой и нулевой фазой в точке предмета (x_0, y_0). Функция пропускания зависит, конечно, и от длины волны λ, однако эта зависимость учитываться не будет, поскольку мы рассматриваем только монохроматический свет.

Пусть $U_0(x_0, y_0)$ — комплексное возмущение в точке, находящейся в плоскости предмета. Элемент поверхности, содержащий точку (x_0, y_0), вызывает возмущение $dU_1(x_1, y_1) = U_0(x_0, y_0) K(x_0, y_0; x_1, y_1) dx_0 dy_0$ в точке (x_1, y_1) плоскости изображения. Следовательно, полное возмущение в точке (x_1, y_1) равно

$$U_1(x_1, y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_0(x_0, y_0) K(x_0, y_0; x_1, y_1)\, dx_0\, dy_0. \qquad (2)$$

Бесконечные пределы интегрирования играют чисто формальную роль, поскольку вне области, которая посылает свет в пространство изображений системы, величина $U_0 K$ равна нулю.

При рассмотрении точечных источников свойства системы описывались комплексным возущением в плоскости выходного зрачка, которое характеризовалось функцией аберраций и амплитудным множителем, причем последний считался постоянным в системах с умеренной апертурой. Функцию пропускания нетрудно выразить через эти величины. Для этого рассмотрим сначала соотношение (2) в предельном случае, когда источник становится точечным, имеет единичную «силу» и нулевую фазу и расположен в точке $x_0 = x_0'$, $y_0 = y_0'$, т. е. когда

$$U_0(x_0, y_0) = \delta(x_0 - x_0')\, \delta(y_0 - y_0'), \qquad (3)$$

где δ — дельта-функция Дирака (см. приложение 4). Тогда (2) дает

$$U_1(x_1, y_1) = K(x_0', y_0'; x_1, y_1), \qquad (4)$$

т. е. функция пропускания K описывает возмущение от точечного источника, задаваемого (3). Выберем центр опорной сферы Гаусса в точке параксиального изображения $x_1' = x_0'$, $y_1' = y_0'$. Пусть R — радиус этой сферы, а

$$H(x_0', y_0'; \xi, \eta) = \frac{i}{\lambda} G(x_0', y_0'; \xi, \eta) \frac{\exp(-ikR)}{R} \qquad (5)$$

— возмущение от точечного источника (3) в ее произвольной точке (ξ, η). Тогда фаза функции G равна (с точностью до аддитивного члена $\pi/2$) функции аберраций Φ системы, а амплитуда G служит мерой неоднородности амплитуды волны, формирующей изображение. Множитель i/λ в правой части (5) введен для упрощения окончательных формул. Согласно принципу Гюйгенса — Френеля возмущение в плоскости изображения связано с возмущением на опорной сфере Гаусса формулой (углы дифракции считаются малыми)

$$U_1(x_1, y_1) = -\frac{i}{\lambda} \iint H(x_0', y_0'; \xi, \eta) \frac{e^{iks}}{s}\, d\xi\, d\eta, \qquad (6)$$

где s — расстояние между точкой (ξ, η) на этой сфере и точкой (x_1, y_1) на плоскости параксиального изображения, а интегрирование проводится по той части

*) (x_0, y_0) и (x_1, y_1) можно считать переменными Зайделя, введенными в § 5.2, если положить $l_1 = C = 1$.

опорной сферы, которая приблизительно закрывает апертуру. Если в уравнениях (8.8.2) и (8.8.7) положить $x=x_1-x_0'$, $y=y_1-y_0'$, $z=0$ и $f=R$, то найдем

$$s\approx R-\frac{(x_1-x_0')\,\xi+(y_1-y_0')\,\eta}{R}. \tag{7}$$

Из формул (4) — (7) получим

$$K(x_0,\ y_0;\ x_1,\ y_1)=$$
$$=\frac{1}{(\lambda R)^2}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}G(x_0,\ y_0;\ \xi,\ \eta)\exp\left\{-\frac{2\pi i}{\lambda R}\left[(x_1-x_0)\xi+(y_1-y_0)\eta\right]\right\}d\xi\,d\eta, \tag{8}$$

где $G=0$ в точках $(\xi,\ \eta)$, находящихся вне отверстия. Это и есть требуемое соотношение между функцией пропускания K и *функцией зрачка* G системы.

Поскольку функцию K можно рассматривать как возмущение в точке изображения точечного источника, то она (рассматриваемая как функция x_1, y_1) имеет резкий максимум в точке параксиального изображения или вблизи от нее $x_1=x_0$, $y_1=y_0$ и быстро спадает, хотя, как правило, немонотонно, с увеличением расстояния до этой точки. В хорошо скоррегированной системе функция K имеет заметную величину лишь в области, размер которой порядка диаметра первого темного кольца в картине Эйри. Как функция координат $(x_0,\ y_0)$ K изменяется очень слабо при перемещении точки по поверхности предмета. Точнее, рабочее поле можно разделить на несколько областей так, чтобы каждая из них была больше самой мелкой детали, которую система способна разрешить, причем в каждой из этих областей A величина K является в хорошем приближении функцией вектора отклонения точки $(x_1,\ y_1)$ от точки параксиального изображения, но не зависит от координат самой точки изображения. Например, в хорошо скоррегированной системе функция $K(x_0,\ y_0;\ x_1,\ y_1)$ описывает с точностью до постоянного множителя картину Эйри, центр которой находится в точке параксиального изображения $(x_0,\ y_0)$. В этом случае можно написать

$$K(x_0,\ y_0;\ x_1,\ y_1)=K_A(x_1-x_0,\ y_1-y_0). \tag{9}$$

Область A, обладающая такими свойствами, называется *изопланатической областью* системы. Ограничимся рассмотрением предметов, малых по сравнению с изопланатической областью *). В этом случае уравнения (2) и (8) можно заменить на

$$U_1(x_1,\ y_1)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}U_0(x_0,\ y_0)\,K(x_1-x_0,\ y_1-y_0)\,dx_0\,dy_0 \tag{2a}$$

и

$$K(x_1-x_0,\ y_1-y_0)=$$
$$=\frac{1}{(\lambda R)^2}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}G(\xi,\ \eta)\exp\left\{-\frac{2\pi i}{\lambda R}\left[(x_1-x_0)\,\xi+(y_1-y_0)\eta\right]\right\}d\xi\,d\eta, \tag{8a}$$

где функция G уже не зависит от координат точки предмета.

---

*) Подробное обсуждение условий, при которых выполняется соотношение (9), изложено в работе [38].

Разложим U_0, U_1 и K в интегралы Фурье:

$$U_0(x_0, y_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{U}_0(f, g) \exp\{-2\pi i [fx_0 + gy_0]\} \, df \, dg, \tag{10а}$$

$$U_1(x_1, y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{U}_1(f, g) \exp\{-2\pi i [fx_1 + gy_1]\} \, df \, dg, \tag{10б}$$

$$K(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{K}(f, g) \exp\{-2\pi i [fx + gy]\} \, df \, dg. \tag{10в}$$

Тогда обратное преобразование Фурье дает

$$\mathcal{U}_0(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_0(x_0, y_0) \exp\{2\pi i [fx_0 + gy_0]\} \, dx_0 \, dy_0, \tag{11а}$$

$$\mathcal{U}_1(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_1(x_1, y_1) \exp\{2\pi i [fx_1 + gy_1]\} \, dx_1 \, dy_1, \tag{11б}$$

$$\mathcal{K}(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(x, y) \exp\{2\pi i [fx + gy]\} \, dx \, dy. \tag{11в}$$

Согласно уравнению (2а) U_1 является сверткой U_0 и K; тогда на основании теоремы о свертке [39] получим после обратного преобразования Фурье простое соотношение

$$\mathcal{U}_1(f, g) = \mathcal{U}_0(f, g) \mathcal{K}(f, g). \tag{12}$$

Отсюда следует, что если возмущения в плоскости предмета и в плоскости изображения представлены в виде суперпозиций пространственных гармоник всех возможных «пространственных частот» f, g, то каждая компонента возмущения в изображении зависит только от соответствующей компоненты в предмете, и их отношение равно $\mathcal{K}$. Таким образом, переход от предмета к изображению эквивалентен действию *линейного фильтра*. Более того, сравнение (10в) с (8а) дает

$$\mathcal{K}\left(\frac{\xi}{\lambda R}, \frac{\eta}{\lambda R}\right) = G(\xi, \eta); \tag{13}$$

т. е. *при когерентном освещении функция частотного отклика* (называемая также *коэффициентом пропускания) $\mathcal{K}(f, g)$ равна значению функции зрачка G в точке*

$$\xi = \lambda R f, \qquad \eta = \lambda R g \tag{14}$$

опорной сферы Гаусса.

Поскольку G равна нулю в точках плоскости ξ, η, лежащих вне отверстия, амплитуды разложения, которые соответствуют частотам, превышающим определенное значение, не пропускаются системой. Если отверстие имеет вид окружности радиуса a, то, очевидно, амплитуды, соответствующие частотам, для которых

$$f^2 + g^2 > \left(\frac{a}{\lambda R}\right)^2, \tag{15}$$

не пропускаются системой. Чтобы проиллюстрировать этот результат, рассмотрим одномерный предмет, свойства которого не изменяются в направлении x. Пусть Δy_0 — период, соответствующий частоте g. Тогда из (15) следует, что система может пропускать информацию только о тех спектральных компонентах, для которых

$$\Delta y_0 = \frac{1}{g} > \frac{\lambda}{\sin\theta_1}, \tag{16}$$

где $\sin\theta_1 \approx a/R$ — половина угловой апертуры со стороны изображения, ко-

торая считается малой. Тогда $y_0 = MY_0$, где M — линейное увеличение, и если предположить, что система подчиняется условию синусов, то (см. п. 4.5.1) $n_0 \sin\theta_0 / n_1 \sin\theta_1 = M$, и (16) можно переписать в виде

$$\Delta Y_0 > \frac{\lambda_0}{n_0 \sin\theta_0}, \tag{17}$$

где $\lambda_0 = n_1\lambda$ — длина волны в вакууме, а $n_0 \sin\theta_0$ — числовая апертура системы. Таким образом, если возмущение в плоскости предмета меняется синусоидально в зависимости от положения, то информация о нем в плоскости изображения появится лишь в том случае, если период возмущения превышает значение правой части (17).

9.5.2. Некогерентное освещение. Рассмотрим теперь случай, когда свет, испускаемый различными участками плоскости предмета, некогерентен, например, когда предмет служит первичным источником. Пусть $I_0(x_0, y_0)$ — интенсивность в произвольной точке плоскости предмета (здесь используются те же координаты, что и раньше). Интенсивность света, достигающего точки (x_1, y_1) в плоскости изображения и выходящего из элемента $dx_0 dy_0$ с центром в точке предмета (x_0, y_0), равна $dI_1(x_1, y_1) = I_0(x_0, y_0)\,|K(x_0, y_0; x_1, y_1)|^2 dx_0 dy_0$, где K — снова функция пропускания системы. Поскольку предполагается, что свет, исходящий из предмета, некогерентен, интенсивности, обусловленные различными элементами плоскости предмета, складываются, так что полная интенсивность в точке (x_1, y_1) равна

$$I(x_1, y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_0(x_0; y_0)\,|K(x_0, y_0; x_1, y_1)|^2 dx_0\,dy_0. \tag{18}$$

Если снова ограничиться рассмотрением достаточно малых предметов, то можно заменить (18) на

$$I_1(x_1, y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_0(x_0, y_0)\,|K(x_1 - x_0, y_1 - y_0)|^2 dx_0\,dy_0. \tag{19}$$

Отметим, что замена (18) на (19) не обязательно требует, чтобы функция K удовлетворяла полному условию изопланатичности (9); достаточно, чтобы только модуль K удовлетворял ему, т. е. чтобы во всей области A, занятой предметом, выполнялось с хорошей точностью соотношение

$$|K(x_0, y_0; x_1, y_1)| = |K_A(x_1 - x_0, y_1 - y_0)|.$$

Исследование Дюмонте [38] показывает, что это условие, как правило, выполняется довольно точно в значительно большей области плоскости предмета, чем определяемой соотношением (9). Следовательно, заменой оптической системы линейным фильтром можно пользоваться шире для некогерентного, чем для когерентного освещения. Однако для нахождения связи функции частотного отклика с функцией зрачка системы в виде относительно простой формулы (см. ниже формулу (22)) мы ограничимся рассмотрением предметов, размеры которых настолько малы, что выполняется полное условие изопланатичности.

Соотношение (19) показывает, что при некогерентном освещении распределение интенсивности в изображении является сверткой распределения интенсивности в предмете, в которое входит квадрат модуля функции пропускания. Представим эти функции в виде интегралов Фурье (10) и обозначим их «пространственные спектры» через $\mathcal{Y}_0(f, g)$, $\mathcal{Y}_1(f, g)$ и $\mathcal{L}(f, g)$. Тогда с помощью обратного преобразования Фурье получим вместо (11) соотношения

$$\mathcal{Y}_0(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_0(x_0, y_0) \exp\{2\pi i\,[fx_0 + gy_0]\}\,dx_0\,dy_0, \tag{20а}$$

$$\mathcal{Y}_1(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_1(x_1, y_1) \exp\{2\pi i\,[fx_1 + gy_1]\}\,dx_1\,dy_1, \tag{20б}$$

$$\mathcal{L}(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |K(x, y)|^2 \exp\{2\pi i\,[fx + gy]\}\,dx\,dy. \tag{20в}$$

Из (19) на основании теоремы о свертке имеем

$$\mathcal{Y}_1(f, g) = \mathcal{Y}_0(f, g)\,\mathcal{L}(f, g). \qquad (21)$$

Таким образом, переход от предмета к изображению снова эквивалентен действию *линейного фильтра*, однако в этом случае преобразованию подвергается не пространственный фурье-образ комплексной амплитуды, а фурье-образ интенсивности. Функцией частотного отклика системы служит функция $\mathcal{L}(f, g)$, которую с помощью (20в), (10в) и теоремы о свертке можно записать в виде

$$\mathcal{L}(f, g) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{K}(f'+f, g'+g)\,\mathcal{K}^*(f', g')\,df'\,dg'. \qquad (22)$$

Интеграл в правой части (22) называется *автокорреляционной функцией* (функции $\mathcal{K}$); он часто встречается при исследовании многих физических задач статистического характера. Мы с ним еще будем иметь дело позже при изложении частичной когерентности.

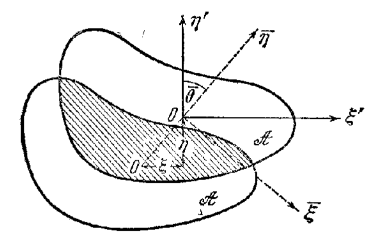


Рис 9.12. Область интегрирования (заштрихована), относящаяся к вычислению функции частного отклика $\mathcal{L}(f,g)$ при некогеречтном освещении для пары частот $f=\xi/\lambda R$ и $g=\eta/\lambda R$.

Ранее было показано, что $\mathcal{K}(f, g)$ есть значение функции зрачка G в соответствующей точке опорной сферы Гаусса. Подставив (13) в (22), получим

$$\mathcal{L}\left(\frac{\xi}{\lambda R}, \frac{\eta}{\lambda R}\right) = \frac{1}{(\lambda R)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(\xi'+\xi, \eta'+\eta)\,G^*(\xi', \eta')\,d\xi'\,d\eta', \qquad (23)$$

таким образом мы установили, что *при некогерентном освещении функция частотного отклика системы* $\mathcal{L}(f, g)$ равна *с точностью до постоянного множителя автокорреляционной функции зрачка системы*. Пусть $\mathcal{A}$ — площадь выходного зрачка. Поскольку функция зрачка $G(\xi', \eta')$ равна нулю за границами отверстия, то площадь участка плоскости ξ', η', на котором подынтегральное выражение в (23) не обращается в нуль, равна площади области, общей для отверстия $\mathcal{A}$ и такого же отверстия, но смещенного относительно первого на расстояния ξ и η в направлении отрицательных ξ' и η' (см. рис. 9.12). Когда величины ξ и η настолько велики, что эти две области не перекрываются, функция частотного отклика равна, очевидно, нулю; таким образом, как и при когерентном освещении, система пропускает лишь те пространственные гармоники, частота которых не превышает определенной максимальной величины. В частности, в случае отверстия в виде круга радиуса a эти области не перекрываются, если $\xi^2+\eta^2 \geqslant (2a)^2$ или если

$$f^2+g^2 > \left(\frac{2a}{\lambda R}\right)^2. \qquad (24)$$

Проводя те же рассуждения, что и при выводе (17) из (15), нетрудно показать, что при использовании некогерентного освещения апланатическая система может пропускать информацию только о таких компонентах, для периода которых справедливо неравенство

$$\Delta Y_0 > \frac{0{,}5\lambda_0}{n_0 \sin\theta_0}. \qquad (25)$$

Как мы видим, предельное значение в этом случае точно равно половине соответствующей величины, получающейся при когерентном освещении.

Хотя функция частотного отклика $\mathcal{L}$ системы зависит от двух переменных f и g, в принципе можно получить полную информацию о ней с помощью экспериментов с одномерными пробными предметами. Рассмотрим для этого пару

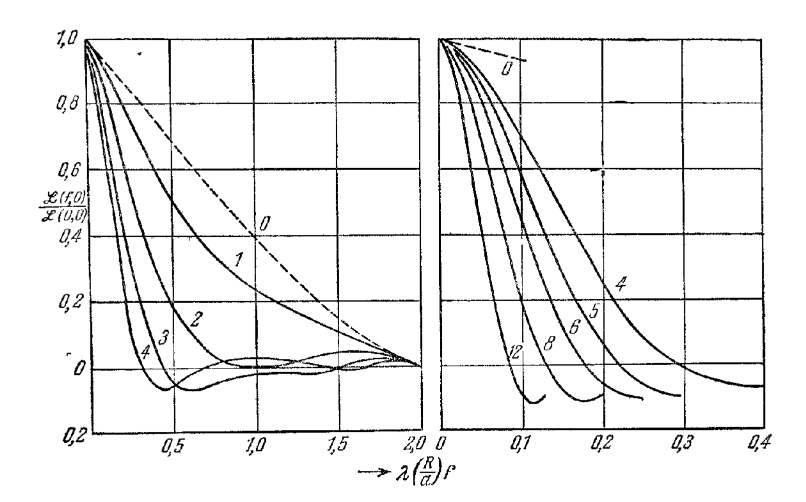


Рис. 9.13. Нормированные кривые частотного отклика при некогерентном освещении несколько расфокусированной системы, свободной от геометрических аберраций [35].

$\Phi=\frac{m\lambda}{\pi}\rho^2$, $|G|=1$. Числа у кривых указывают значения параметра $m=\frac{\pi}{2\lambda}\left(\frac{a}{R}\right)^2 z$, где z — расстояние между плоскостью наблюдения и параксиальной фокальной плоскостью

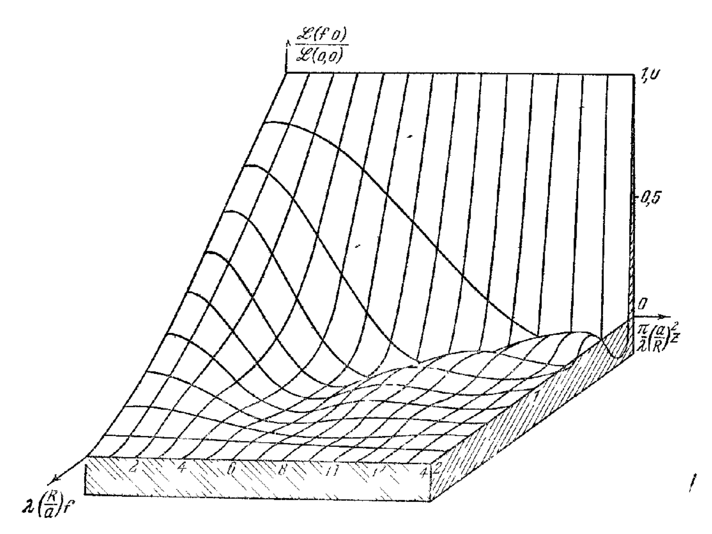


Рис. 9.14. Нормированные кривые частотного отклика при некогерентном освещении несколько расфокусированной системы, свободной от геометрических аберраций, в зависимости от пространственной частоты f и расфокусировки z, $|G| = 1$ [40].

Кривые на рис. 9.13 являются сечениями данной поверхности плоскостями, перпендикулярными к оси z.

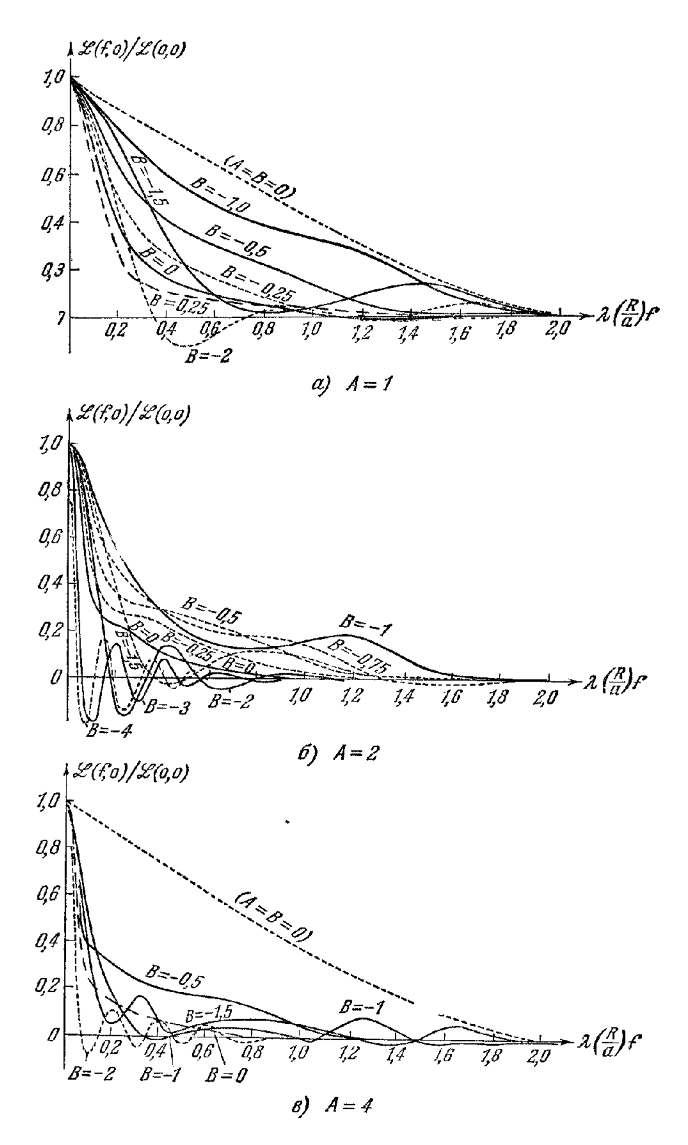


Рис. 9.15. Нормированные кривые частотного отклика при некогерентном освещении и выбранных положениях фокуса системы, обладающей небольшой первичной сферической аберрацией $\Phi = A(\rho^4 + B\rho^2)\lambda$, $|G| = 1$ [41].

Значение $B=0$ соответствует параксиальной фокальной плоскости, $B=-2$ – плоскости изображения, проходящей через краевой фокус.

частот (f, g) и введем полярные координаты в пространстве частот $f = h \sin\theta$, $g = h\cos\theta$. Предположим теперь, что оси повернулись в своей плоскости на угол $\overline{\theta}$ в положительном направлении θ. Тогда f и g преобразуются в $\overline{f} = h \sin(\theta - \overline{\theta})$ и $\overline{g} = h\cos(\theta - \overline{\theta})$, но величина $\mathscr{L}$ остается, очевидно, неизменной. Мы можем выбрать угол поворота $\overline{\theta}$ равным $\operatorname{arctg}(f/g)$, при котором новая ось $\eta(O\overline{\eta})$ будет лежать вдоль линии OO (см. рис. 9.12). Тогда $\overline{f} = 0$, $\overline{g} = \sqrt{f^2+g^2}$, и можно заключить, что *значение функции частотного отклика $\mathscr{L}$ оптической системы для пары частот (f, g) равно значению функции частотного отклика одномерной структуры с частотой $\sqrt{f^2+g^2}$ и направлением периодичности, составляющим угол $\operatorname{arctg}(f/g)$ с меридиональной плоскостью $\theta = 0$.* Этот результат существенно упрощает получение аналитических выражений для функции частотного отклика для любой системы. Аналогичный результат справедлив, конечно, и для функции частотного отклика $\mathscr{K}(f, g)$ в случае когерентного освещения, однако он имеет значительно меньшее практическое значение.

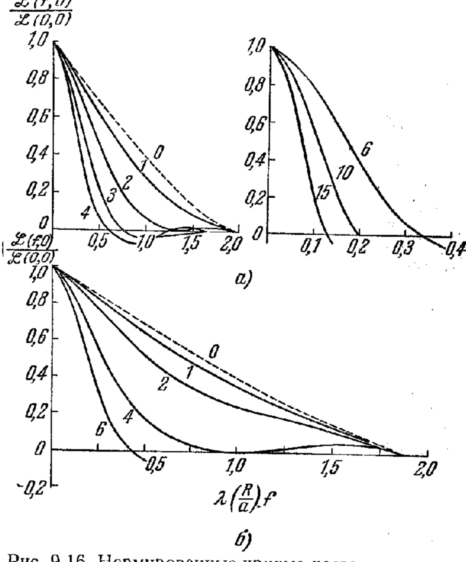


Рис. 9.16. Нормированные кривые частотного отклика для некогерентного освещения при наличии первичного астигматизма $\Phi = \frac{m\lambda}{\pi}\rho^2\cos^2\theta$, $|G| = 1$ [42].

Плоскость изображения расположена посередине между тангенциальной и сагиттальной фокальными линиями. Линии на *а* периодичны вдоль меридиана $\theta = \pi/4$, а на *б* — вдоль меридианов $\theta = 0$ или $\theta = \pi/2$. Числа на кривых указывают значения параметра m

Рассмотрим теперь частотный отклик свободной от аберраций, но слегка расфокусированной центрированной системы. Из обсуждения, изложенного в п. 9.1.2, следует, что смещение плоскости изображения на небольшое расстояние z в положительном направлении оси z формально эквивалентно введению волновой аберрации, равной

$$\Phi(\xi, \eta) = \frac{1}{2}\left(\frac{a}{R}\right)^2 z\rho^2, \quad \left(\rho^2 = \frac{\xi^2+\eta^2}{a^2} \leqslant 1\right); \tag{26}$$

таким образом, если на опорной сфере Гаусса амплитуда волны постоянна, то с точностью до постоянного множителя функция зрачка запишется в виде

$$G(\xi, \eta) = \exp[ik\Phi(\xi, \eta)] = \exp\left[\frac{i}{2}\left(\frac{a}{R}\right)^2 z\rho^2\right]. \tag{27}$$

Функцию частотного отклика можно найти из (23) и (27); полученные таким способом графики представлены на рис. 9.13 и 9.14. Из них видно, что при больших частотах отклик системы быстро спадает, если смещение фокуса превышает значение, соответствующее $\Phi = \frac{\lambda}{\pi}\rho^2$, т. е. значение $z = \frac{2\lambda}{\pi}\left(\frac{R}{a}\right)^2$.

На рис. 9.15 и 9.16 показаны кривые частотного отклика для систем с первичной сферической аберрацией и первичным астигматизмом

Обзор технических средств, использующихся для измерения частотного отклика, приведен в [43].

ЛИТЕРАТУРА

1. Rayleigh, Phil. Mag. 8, 403 (1879).
2. E. Wolf, в сб. «Rep. Progr. Phys.», Phys. Soc., 1951, Vol. 14, p. 95.
3. G. C. Steward, Phil. Trans. Roy. Soc. A225, 131 (1925).
4. G. C. Steward, The Symmetrical Optical System, Cambr. Univ. Press, 1928.
5. J. Picht, Ann. d. Physik 77, 685 (1925).
6. J. Picht, Ann. d. Physik 80, 491 (1926).
7. J. Picht, Optische Abbildung, Vieweg, Braunschweig, 1931.
8. M. Born, Naturwissensch. 20, 921 (1932).
9. M. Born, Optik, Springer Berlin 1933, p. 202. (М. Борн, Оптика, ГНТИУ Харьков — Киев, 1937.)
10. B. R. A. Nijboer, Thesis, University of Groningen, 1942.
11. B. R. A. Nijboer, Physica 10, 679 (1943).
12. B. R. A. Nijboer, Physica 13, 605 (1947).
13. F. Zernike, B. R. A. Nijboer, в сб. «La Théorie des Images Optiques», Rev. d'Opt., Paris, 1949, p. 227.
14. K. Nienhuis, B. R. A. Nijboer, Physica 14, 590 (1949).
15. K. Nienhuis, Thesis, University of Groningen, 1948.
16. M. P. Bachynski, G. Bekefi, Trans. IRE AP-4, 412 (1956).
17. N. G. van Kampen, Physica 14, 575 (1949).
18. N. G. van Kampen, Physica 16, 817 (1950).
19. N. G. van Kampen, Physica 24, 437 (1958).
20. J. Focke, Optica Acta 3, 110 (1956).
21. K. Strehl, Z. f. Instrumkde 22, 213 (1902).
22. R. Courant, D. Hilbert, Methods of Mathematical Physics, Intersci. Publ., 1953, vol. 1. (Р. Курант, Д. Гильберт, Методы математической физики, Гостехиздат, М.—Л., 1951, т. 1.)
23. Zernike, Physica 1, 689 (1934).
24. R. Richter, Z. f. Instrumkde 45, 1 (1925).
25. A. Maréchal, Rev. d'Opt. 26, 257 (1947).
26. G. N. Watson, A Treatise on the Theory of Bessel Functions, Cambr. Univ. Press, 2nd ed., 1944, p. 368. (Г. Ватсон, Теория бесселевых функций, ИЛ, 1949.)
27. E. T. Whittaker, G. N. Watson, A Course of Modern Analysis, Cambr. Univ. Press, 4th ed. 1940. (Э. Уиттекер и Г. Ватсон, Курс современного анализа, Физматгиз, 1963.)
28. E. H. Linfoot, Recent Advances in Optics, Clarendon Press, Oxford, 1955 pp. 60—61.
29. M. Françon, в кн. «Encyclopedia of Physics», ed. S. Flügge, Springer, Berlin, 1956, Vol. 24, pp. 321—322.
30. R. Kingslake, Proc. Phys. Soc. 61, 147 (1948).
31. P. M. Duffieux, L'Intégrale de Fourier et ses Applications a l'Optique, Rennes, 1946.
32. P. M. Duffieux, G. Lansraux, Rev. d'Opt. 24, 65, 151, 215 (1945).
33. A. Blanc-Lapierre, Ann. de l'Inst. Henri Poincaré 13, 245 (1953).
34. H. H. Hopkins, Proc. Roy. Soc. A217, 408 (1953).
35. H. H. Hopkins, Proc. Roy. Soc. A231, 91 (1955)
36. H. H. Hopkins, Proc. Phys. Soc. B69, 562 (1956).
37. K. Miyamoto, в сб.: «Progr. in Optics», vol. 1, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1961, p. 41.
38. P. Dumontet, Optica Acta 2, 53 (1955).
39. I. N. Sneddon, Fourier Transforms, McGrow-Hill, New York, 1951, p. 23.
40. W. H. Steel, Optica Acta 3, 67 (1956).
41. G. Black, E. H. Linfoot, Proc. Roy. Soc. A239, 522 (1957).
42. M. De, Proc. Roy. Soc. A233, 96 (1955).
43. K. Murata, в сб.: «Progr. in Optics», vol. 5, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1965, p. 199.

ГЛАВА 10

ИНТЕРФЕРЕНЦИЯ И ДИФРАКЦИЯ ЧАСТИЧНО КОГЕРЕНТНОГО СВЕТА

§ 10.1. Введение

До сих пор мы имели дело главным образом с монохроматическим светом, излучаемым точечным источником. Свет от реального физического источника никогда не бывает строго монохроматическим, так как даже самая узкая спектральная линия обладает конечной шириной. Кроме того, физический источник имеет конечные размеры и состоит из огромного числа элементарных излучателей (атомов). Согласно теореме Фурье возмущение, создаваемое таким источником, можно выразить в виде суммы строго монохроматических и поэтому бесконечно длинных волновых цугов. В элементарной теории, оперирующей с монохроматическим светом, по существу рассматривается одна компонента такого фурье-представления.

В монохроматическом волновом поле амплитуда колебаний в любой точке P постоянна, тогда как фаза линейно меняется со временем. В отличие от этого в волновом поле, создаваемом реальным источником, амплитуда и фаза претерпевают нерегулярные флуктуации, частота которых существенно зависит от эффективной ширины спектра $\Delta\nu$. Комплексная амплитуда остается более или менее постоянной лишь в течение промежутка времени Δt, малого по сравнению с величиной, обратной эффективной ширине спектра $\Delta\nu$. Изменение разности фаз любых двух фурье-компонент за этот промежуток значительно меньше 2π, и поэтому сумма таких компонент представляет собой возмущение, которое в течение указанного промежутка времени ведет себя подобно монохроматической волне со средней частотой. Однако для более длинного интервала времени это несправедливо. Характеристическое время $\Delta t = 1/\Delta\nu$ равно по порядку величины *времени когерентности*, введенному в п. 7.5.8.

Рассмотрим световые возмущения в двух точках P_1 и P_2 в волновом поле, создаваемом протяженным квазимонохроматическим источником. Для простоты предположим, что волновое поле создается в вакууме, а расстояние от источника до точек P_1 и P_2 составляет много длин волн. Можно ожидать, что при достаточной близости P_1 и P_2 друг к другу флуктуации амплитуд и фаз возмущения в этих точках не будут независимыми. Если P_1 и P_2 так близки друг к другу, что разность между расстояниями до любой точки S источника $\Delta\mathcal{S} = SP_1 - SP_2$ мала по сравнению со средней длиной волны $\bar{\lambda}$, то разумно предположить, что флуктуации амплитуд и фаз в P_1 и P_2 одинаковы. Можно также считать, что между этими флуктуациями будет существовать некоторая корреляция даже при большем удалении P_1 от P_2 при условии, что для всех точек источника разность расстояний $\Delta\mathcal{S}$ не превышает длину когерентности $c\Delta t \sim c/\Delta\nu = \bar{\lambda}^2/\Delta\lambda$. Таким образом, мы приходим к понятию *области когерентности* вокруг любой точки P волнового поля.

Для адекватного описания волнового поля, создаваемого конечным полихроматическим источником, желательно ввести некоторую меру корреляции, существующей между колебаниями в различных точках P_1 и P_2 поля. Естественно полагать, что такая мера будет тесно связана с резкостью интерференционных полос, которые получились бы при сложении колебаний в этих точках. Нужно ожидать резких полос при сильной корреляции (например, когда свет приходит в P_1 и P_2 из очень маленького источника, испускающего излучение

в узком спектральном интервале) и полного отсутствия полос в отсутствие корреляции (например, когда каждая из точек P_1 и P_2 получает свет от различных физических источников). Для описания таких ситуаций мы использовали соответственно термины «когерентный» и «некогерентный». В общем случае не реализуется ни одна из этих ситуаций, и нужно говорить о *частично когерентных* колебаниях.

По-видимому, первые исследования, относящиеся к вопросу о частичной когерентности, были выполнены Верде [1], который изучал размеры области когерентности для света от протяженного первичного источника. Позже в исследованиях Майкельсона была установлена связь между видностью интерференционных полос и распределением яркости по поверхности протяженного первичного источника [2] (см. п. 7.3.6), а также между видностью и распределением энергии в спектральной линии [3] (см. п. 7.5.8). Фактически результаты Майкельсона были интерпретированы на языке корреляций лишь значительно позднее, однако его исследования внесли существенный вклад в формулировку современных теорий частичной когерентности (см. [4]). Первую количественную меру корреляции световых колебаний ввел Лауэ [5] при исследованиях по термодинамике световых пучков. Дальнейший вклад в теорию был внесен Береком *) [6], который использовал понятие корреляции при исследовании образования изображения в микроскопе.

Новый этап в развитии этого вопроса начался после опубликования статьи Ван-Циттерта [8], который определил совместное распределение вероятностей для световых возмущений в двух любых точках экрана, освещенного протяженным первичным источником. В следующей статье [9] (см. также [10]) он определил распределение вероятностей для световых возмущений в одной произвольной точке, но в два различных момента времени. В пределах точности его расчетов распределения вероятностей оказались гауссовыми, и он определил соответствующие коэффициенты корреляции.

Новый, более простой подход к проблеме частичной когерентности был предложен Цернике в важной работе [11], опубликованной в 1938 г. Его определение «степени когерентности» световых колебаний прямо связано с экспериментом. Он получил также ряд ценных результатов, относящихся к этой величине. Хотя степень когерентности, введенная Цернике, для большинства практических случаев эквивалентна коэффициенту корреляции Ван-Циттерта и близка к аналогичной величине, предложенной Лауэ, его методы, по-видимому, особенно хороши для решения практических задач инструментальной оптики. Гопкинс [12] значительно упростил эти методы и применил их к изучению формирования изображения и разрешающей силы **).

Перечисленные выше исследования связали два крайних случая, а именно случаи полной когерентности или полной некогерентности. Однако полученные результаты страдали некоторой ограниченностью, так как они относились главным образом к квазимонохроматическому свету и к случаям с достаточно малой разностью хода интерферирующих пучков. Для рассмотрения более сложных ситуаций и формулировки теории на строгой основе было необходимо дальнейшее обобщение. Оно было выполнено Вольфом [15, 16] и независимо Блан-Лапьерром и Дюмонте [17]; в этих работах были введены более общие корреляционные функции. Оказалось, что такие функции строго удовлетворяют двум волновым уравнениям; отсюда следует, что не только оптическое возмущение, но и корреляция между возмущениями распространяются в форме волн. В свете этого вывода многие из ранее выведенных теорем получают относительно простое истолкование.

---

*) Эксперименты, относящиеся к исследованиям Берека, описаны в [7].

**) В сходных исследованиях Д. Габор и Г. Гамо использовали понятие частичной когерентности при рассмотрении оптической передачи в рамках теории информации [13, 14].

Названные выше корреляционные функции характеризуют корреляцию между световыми колебаниями в *двух* пространственно-временных точках. Такие корреляционные функции «второго порядка» позволяют провести полный анализ обычных оптических экспериментов, в том числе по интерференции и дифракции света от постоянных источников *). Аналогичные корреляционные функции можно применить и для анализа нестационарных полей, но тогда эти функции необходимо определять через величины, усредненные по ансамблю, а не по времени. (Для стационарного поля усреднения обоих типов обычно приводят к одинаковому результату.) Однако мы ограничимся обсуждением лишь стационарных полей, поскольку анализ нестационарных полей требует значительного усложнения математического аппарата и пока нет никаких экспериментов, относящихся к эффектам когерентности при нестационарных световых полях. Для анализа более сложных экспериментов могут потребоваться корреляционные функции более высокого порядка, т. е. такие, для которых сумма степеней величин, характеризующих поле, больше двух. Однако до сих пор подобные корреляционные функции практически не использовались.

Когда свет излучается тепловым источником, например раскаленным веществом или газовым разрядом, мы можем предположить, что совместное распределение вероятностей для поля в n пространственно-временных точках с хорошим приближением будет гауссовым. Как хорошо известно (см., например, [18]), такие распределения полностью определяются корреляционными функциями второго порядка. Это означает, что для света от теплового источника все корреляционные функции более высокого порядка можно выразить через корреляционные функции второго порядка. Однако для света нетепловых источников, например лазера, последнее несправедливо.

Ряд эффектов когерентности более высокого порядка и соответствующие корреляционные функции кратко рассмотрены в работах [19, 20] (см. также работу Вольфа ([21], стр. 29), где проведена систематическая классификация эффектов когерентности).

Глаубер [22] ввел аналогичные квантовомеханические корреляционные функции, а Сударшан [23] рассмотрел связь между классическим и квантовым описаниями (см. также [24], где содержится обзор эффектов когерентности второго и более высокого порядков).

Теория частичной когерентности привлекательна тем, что она оперирует с величинами (а именно с корреляционными функциями и с усредненными по времени интенсивностями), которые в принципе можно определить из эксперимента. В этом она является полной противоположностью элементарной оптической волновой теории, где из-за очень большой частоты оптических колебаний основную величину невозможно измерить. В настоящей главе мы будем исследовать свойства частично когерентных волновых полей и проиллюстрируем результаты рядом примеров, представляющих практический интерес. Мы коснемся лишь случая светового поля, но наш анализ годится и для других полей. В частности, подобный подход может использоваться в связи с корреляционной методикой, применяемой для изучения радиозвезд [25] и для исследования ионосферы с помощью радиоволн [26].

Математический аппарат, используемый при рассмотрении частичной когерентности, применим также для анализа частичной поляризации. Здесь мы коснемся явлений, которые можно интерпретировать через корреляцию между ортогональными компонентами электромагнитных векторов поля. Первые исследования в этом направлении проведены Дж. Дж. Стоксом [27] (см. также [28]). Современные теории, в которых применяются понятия корреляционных функций и корреляционных матриц, развиты главным образом Винером [29, 30], Перреном [31], Вольфом [16, 32, 33] и Панчаратнамом [34]. Эта проблема будет рассмотрена в заключительном разделе (см. § 10.8) настоящей главы **).

*) Точнее, от источников, создающих стационарное поле, определенное ниже, на стр 460.

**) Новые результаты, относящиеся к частичной когерентности, изложены в [89*, 90*] (см. также [91*—93*]). (*Прим. перев.*)

§ 10.2. Комплексное представление вещественных полихроматических полей

Изучая монохроматические волновые поля, мы установили, что полезно рассматривать каждую вещественную волновую функцию как вещественную часть соответствующей комплексной волновой функции. В настоящей главе мы займемся полихроматическими (т. е. немонохроматическими) полями. Здесь также полезно использовать комплексное представление, которое можно считать естественным обобщением представления, применявшегося для монохроматических полей.

Пусть $V^{(r)}(t)$ $(-\infty \leqslant t \leqslant \infty)$ — вещественное возмущение, например декартова компонента электрического вектора, в фиксированной точке пространства. Предположим, что $V^{(r)}(t)$ квадратично интегрируемо. Его можно выразить в виде интеграла Фурье

$$V^{(r)}(t) = \int_0^\infty a(\nu) \cos[\varphi(\nu) - 2\pi\nu t]\, d\nu. \tag{1}$$

Свяжем с $V^{(r)}$ комплексную функцию

$$V(t) = \int_0^\infty a(\nu) \exp\{i[\varphi(\nu) - 2\pi\nu t]\}\, d\nu. \tag{2}$$

Тогда имеем

$$V(t) = V^{(r)}(t) + iV^{(i)}(t), \tag{3}$$

где

$$V^{(i)}(t) = \int_0^\infty a(\nu) \sin[\varphi(\nu) - 2\pi\nu t]\, d\nu. \tag{4}$$

Функции $V^{(i)}(t)$ и $V(t)$ однозначно определяются функцией $V^{(r)}(t)$, поскольку $V^{(i)}(t)$ получается из $V^{(r)}(t)$ при замене фазы $\varphi(\nu)$ каждой фурье-компоненты на $\varphi(\nu) - \pi/2$. Интегралы (1) и (4) называют *сопряженными интегралами Фурье* или *сопряженными функциями*. Можно показать *), что они получаются друг из друга с помощью преобразований Гильберта, т. е.

$$V^{(i)}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{V^{(r)}(t')}{t'-t}\, dt', \qquad V^{(r)}(t) = -\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{V^{(i)}(t')}{t'-t}\, dt', \tag{5}$$

где P — главное значение интеграла по Коши при $t' = t$.

Таким комплексным представлением часто пользуются в теории связи, где V называют *аналитическим сигналом* **), связанным с $V^{(r)}$. Он получил это название потому, что при $V^{(r)}$, удовлетворяющем определенным общим условиям непрерывности, функции $V(z)$, рассматриваемая как функция комплексной переменной z, аналитична в нижней полуплоскости z (см. [35]).

Для дальнейшего укажем переход от $V^{(r)}$ к V, когда $V^{(r)}$ представлено интегралом Фурье вида

$$V^{(r)}(t) = \int_{-\infty}^{\infty} v(\nu) \exp(-2\pi i t\nu)\, d\nu. \tag{6}$$

Так как функция $V^{(r)}$ вещественна, то

$$v(-\nu) = v^*(\nu). \tag{7}$$

---

*) См., например, [35], гл. 5.

**) Понятие аналитического сигнала было введено в работе [36] (см. также [37—39]). Комплексные функции действительной переменной, вещественная и мнимая части которых связаны преобразованиями Гильберта, играют важную роль во многих разделах физики и техники. В физике преобразования Гильберта часто называют *дисперсионными соотношениями*, так как впервые они появились в теории дисперсии света, вызываемой атомами [40] (см. также [41]).

Используя последнее соотношение, мы можем переписать (6) в форме (1) и после сравнения получим

$$v(\nu)=\frac{1}{2}a(\nu)\exp[i\varphi(\nu)], \qquad \nu\geqslant 0. \tag{8}$$

Интеграл (2), выраженный через v, запишется в виде

$$V(t)=2\int_0^\infty v(\nu)\exp(-2\pi i\nu t)\,d\nu. \tag{9}$$

Следовательно, $V(t)$ можно вывести из $V^{(r)}(t)$, представляя $V^{(r)}$ как интеграл Фурье в виде (6), пренебрегая амплитудами, связанными с отрицательными частотами, и удваивая амплитуды, связанные с положительными частотами. По этой причине функцию V называют также связанной с $V^{(r)}$ *комплексной функцией, спектр Фурье которой не содержит отрицательных частот.* Очевидно также, что если спектр Фурье комплексной функции V не содержит амплитуд, связанных с отрицательными частотами, то вещественная и мнимая части V являются сопряженными функциями. Отметим следующие соотношения, которые вытекают из (6), (7) и (9) на основании теоремы Парсеваля и соотношения (3):

$$\int_{-\infty}^{\infty}(V^{(r)}(t))^2\,dt=\int_{-\infty}^{\infty}(V^{(i)}(t))^2\,dt=\frac{1}{2}\int_{-\infty}^{\infty}V(t)V^*(t)\,dt=$$
$$=\int_{-\infty}^{\infty}|v(\nu)|^2d\nu=2\int_0^\infty|v(\nu)|^2d\nu. \tag{10}$$

В большинстве рассматриваемых нами приложений спектральные амплитуды заметно отличаются от нуля лишь в частотном интервале шириной $\Delta\nu$, малом по сравнению со средней частотой $\bar{\nu}$. В этом случае аналитический сигнал допускает простую интерпретацию. Запишем V в виде

$$V(t)=A(t)\exp\{i[\Phi(t)-2\pi\bar{\nu}t]\}, \tag{11}$$

где $A\,(\geqslant 0)$ и Φ вещественны. Согласно (9) и (11)

$$A(t)\exp[i\Phi(t)]=2\int_0^\infty v(\nu)\exp[-2\pi i(\nu-\bar{\nu})t]\,d\nu=\int_{-\bar{\nu}}^{\infty}g(\mu)\exp[-2\pi i\mu t]\,d\mu, \tag{12}$$

где

$$g(\mu)=2v(\bar{\nu}+\mu). \tag{13}$$

По предположению, спектральные амплитуды заметно отличаются от нуля только вблизи $\nu=\bar{\nu}$, и поэтому $|g(\mu)|$ будет заметной величиной лишь около $\mu=0$. Следовательно, интеграл (12) представляет собой суперпозицию гармоник низких частот, а так как $\Delta\nu/\bar{\nu}\ll 1$, то $A(t)$ и $\Phi(t)$ будут медленно меняющимися *) (по сравнению с $\cos 2\pi\bar{\nu}t$ и $\sin 2\pi\bar{\nu}t$) функциями t. Выразим $V^{(r)}$ и $V^{(i)}$ — вещественную и мнимую части V — через A и Φ:

$$\left.\begin{aligned}V^{(r)}(t)&=A(t)\cos[\Phi(t)-2\pi\bar{\nu}t],\\ V^{(i)}(t)&=A(t)\sin[\Phi(t)-2\pi\bar{\nu}t].\end{aligned}\right\} \tag{14}$$

*) Очевидно, что, согласно (14), при этих предположениях мы можем написать

$$V^{(i)}(t)\sim V^{(r)}\left(t+\frac{1}{4\bar{\nu}}\right).$$

В этих формулах $V^{(r)}$ и $V^{(i)}$ выражены в виде модулированных сигналов несущей частоты $\bar{\nu}$. Мы видим, что комплексный аналитический сигнал тесно связан с *огибающей* реального сигнала *). Огибающая $A(t)$ и соответствующий фазовый фактор $\Phi(t)$ выражаются через аналитический сигнал V следующим образом:

$$\left.\begin{aligned} A(t) &= \sqrt{(V^{(r)})^2 + (V^{(i)})^2} = \sqrt{VV^*} = |V|, \\ \Phi(t) &= 2\pi\bar{\nu}t + \operatorname{arctg}\frac{V^{(i)}}{V^{(r)}} = 2\pi\bar{\nu}t + \operatorname{arctg}\left(i\frac{V^*-V}{V^*+V}\right). \end{aligned}\right\} \qquad (15)$$

Таким образом, $A(t)$ не зависит от точного выбора $\bar{\nu}$, а зависимость $\Phi(t)$ от $\bar{\nu}$ представлена только аддитивным членом $2\pi\bar{\nu}t$. Конечно, мы могли бы выбрать в (14) вместо $\bar{\nu}$ любую другую частоту $\bar{\nu}'$, не изменяя значения A; выражение для нового фазового множителя отличалось бы от выражения (15) лишь тем, что вместо $\bar{\nu}$ стояло бы $\bar{\nu}'$.

При выводе (14) и (15) мы не пользовались тем, что сигнал узкополосный ($\Delta\nu/\bar{\nu} \ll 1$), так что эти соотношения являются общими. Однако понятие огибающей полезно лишь при условии $\Delta\nu/\bar{\nu} \ll 1$.

Мы предполагали, что «возмущение» $V^{(r)}(t)$ определяется для всех значений t. Практически же возмущение существует лишь в течение конечного интервала времени — $T \leqslant t \leqslant T$, но, как правило, он значительно превышает интервал, имеющий в данном случае физический смысл масштаба времени (средний период $1/\bar{\nu}$ и время когерентности $1/\Delta\nu$); поэтому можно считать, что $T \to \infty$. Такая идеализация желательна с математической точки зрения из-за предположения о стационарности поля (см. п. 10.3.1). Очевидно, в этом случае необходимо также предположить, что средняя по времени интенсивность (пропорциональная $(V^{(r)})^2$) стремится к конечному значению, когда интервал времени, по которому производится усреднение, неограниченно увеличивается, т. е. что

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (V^{(r)}(t))^2\, dt \qquad (16)$$

конечен. Если этот предел конечен и не равен нулю, то ясно, что интеграл $\int_{-\infty}^{\infty} (V^{(r)}(t))^2\, dt$ расходится. Тем не менее и здесь можно воспользоваться аппаратом анализа Фурье **). Определим «обрезанные» функции следующим образом:

$$\left.\begin{aligned} V_T^{(r)}(t) &= V^{(r)}(t), \quad \text{когда} \quad |t| \leqslant T, \\ V_T^{(r)}(t) &= 0, \quad \text{когда} \quad |t| > T. \end{aligned}\right\} \qquad (17)$$

Так как каждую такую обрезанную функцию мы вправе считать интегрируемой с квадратом, ее можно выразить через интеграл Фурье, например в виде

$$V_T^{(r)}(t) = \int_{-\infty}^{\infty} v_T(\nu) \exp(-2\pi i\nu t)\, d\nu. \qquad (18a)$$

*) Свойства аналитических сигналов как огибающих изучались в работе [42].

**) С проблемой анализа функций времени, которые не исчезают при t, стремящемся к бесконечности, столкнулись на пороге XX века физики, занимавшиеся изучением природы белого света и шума (особенно Л. Гуи, лорд Рэлей и А. Шустер). Строгий математический аппарат был развит главным образом Н. Винером в его работе по обобщению гармонического анализа [43], где изложена также история проблемы и приведена обширная библиография.

Пусть $V_T^{(i)}$ — сопряженная функция, а V_T — ассоциированный аналитический сигнал, т. е.

$$V_T(t)=V_T^{(r)}(t)+iV_T^{(i)}(t)=2\int_0^\infty v_T(\nu)\exp(-2\pi i\nu t)\,d\nu. \qquad (18б)$$

При этом соотношения (10) будут выполняться, если $V^{(r)}$ заменить на $V_T^{(r)}$ и т. д. Следовательно, разделив каждое выражение на $2T$, получим *)

$$\frac{1}{2T}\int_{-\infty}^{\infty}(V_T^{(r)}(t))^2\,dt=\frac{1}{2T}\int_{-\infty}^{\infty}(V_T^{(i)}(t))^2\,dt=$$

$$=\frac{1}{2}\frac{1}{2T}\int_{-\infty}^{\infty}V_T(t)V_T^*(t)\,dt=\int_{-\infty}^{\infty}G_T(\nu)\,d\nu=2\int_0^\infty G_T(\nu)\,d\nu, \qquad (19)$$

где

$$G_T(\nu)=\frac{|v_T(\nu)|^2}{2T}. \qquad (20)$$

Казалось бы естественным перейти теперь к пределу $T\to\infty$. К сожалению, во многих случаях, представляющих практический интерес, функция $G_T(\nu)$, известная как *периодограмма*, не стремится к пределу, а флуктуирует с увеличением T (см., например, [44]; см. также [45]). Однако эту трудность можно преодолеть соответствующей процедурой «сглаживания». Например, будем считать, как принято в теории случайных процессов, что функция $V^{(r)}(t)$ является произвольным членом ансамбля функций, характеризующего статистические особенности процесса Тогда при соответствующих предположениях о природе ансамбля (стационарный он или эргодический) **) можно показать, что флуктуации среднего значения $G_T(\nu)$, взятого по ансамблю функций $V^{(r)}(t)$, стремятся к нулю при $T\to\infty$. Таким образом, обозначив среднее по ансамблю чертой, получим для предела «сглаженной периодограммы»

$$\overline{G_T(\nu)}=\frac{\overline{|v_T(\nu)|^2}}{2T} \qquad (21)$$

выражение ***)

$$G(\nu)=\overline{\lim_{T\to\infty}G_T(\nu)}=\lim_{T\to\infty}\frac{\overline{|v_T(\nu)|^2}}{2T}. \qquad (22)$$

Если $\langle\ \rangle$ означает усреднение по времени, т. е.

$$\langle F(t)\rangle=\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty}F_T(t)\,dt, \qquad (23)$$

*) Подвергая преобразованию Гильберта «обрезанную» функцию, мы не обязательно снова получим «обрезанную» функцию; поэтому в общем случае $V_T^{(i)}$ и V_T не равны нулю вне интервала — $T\leqslant t\leqslant T$. По этой причине, а также для того чтобы избежать определенных математических ухищрений, в качестве пределов интегрирования по времени в (19) и (23) берется $\pm\infty$, а не $\pm T$.

**) Эти понятия рассматриваются, например, в [44, 46, 47].

***) Здесь невозможно привести строгое доказательство существования такого предела и вывести некоторые соотношения, введенные эвристически в настоящем разделе, поскольку это завело бы нас слишком далеко в эргодическую теорию (см. [48], гл. 11; см. также [46], § 8.4 или [47], стр. 38—44).

Вместо усреднения по ансамблю можно использовать и другие операции «сглаживания» (см. [46] или [49], стр. 280—284).

Другой способ определения спектра мощности $G(\nu)$, не связанный с понятием ансамбля, указан в подстрочном примечании на стр. 462.

то в предельном случае $T \to \infty$ получим следующие соотношения, аналогичные (19):

$$\langle (V^{(r)}(t))^2 \rangle = \langle (V^{(i)}(t))^2 \rangle = \frac{1}{2} \langle V(t) V^*(t) \rangle = \int_{-\infty}^{\infty} G(\nu)\, d\nu = 2 \int_{0}^{\infty} G(\nu)\, d\nu. \quad (24)$$

В теории стационарных случайных процессов функция $G(\nu)$, определяемая соотношением (22), называется *спектром мощности* случайного процесса, характеризующегося ансамблем функций $V^{(r)}(t)$. Поскольку в наших рассуждениях $V^{(r)}(t)$ представляет световое возмущение, величина $G(\nu)\, d\nu$ пропорциональна вкладу в интенсивность, обусловленному частотами в интервале $(\nu, \nu + d\nu)$. Мы будем называть $G(\nu)$ *спектральной плотностью* световых колебаний.

Так как $V_T^{(r)}$ — вещественная часть V_T, во всех расчетах, в которых производятся линейные операции над $V_T^{(r)}$, можно пользоваться величиной V_T и выделять вещественную часть только в конечном результате. Более того, так же как и в случае монохроматических полей, соотношение $\langle (V^{(r)})^2 \rangle = \frac{1}{2} \langle VV^* \rangle$ позволяет выразить среднее по времени значение квадрата вещественного возмущения непосредственно через комплексное возмущение, которое мы связали с вещественным.

§ 10.3. Корреляционные функции световых пучков

10.3.1. Интерференция двух частично когерентных пучков. Взаимная функция когерентности и комплексная степень когерентности. Для удовлетворительного решения проблем, в которых фигурирует излучение с конечным набором длин волн, испускаемое конечным источником, необходимо, как мы указывали в § 10.1, установить возможную корреляцию между колебаниями в двух произвольных точках волнового поля. Подходящую меру этой корреляции можно предложить, исходя из анализа эксперимента по интерференции двух пучков.

Рассмотрим волновое поле, образованное протяженным полихроматическим источником σ. Вначале пренебрежем эффектами поляризации и будем считать световое возмущение вещественной скалярной функцией положения и времени $V^{(r)}(P, t)$. Функции $V^{(r)}(P, t)$ поставим в соответствие аналитический сигнал $V(P, t)$. Конечно, невозможно наблюдать, как эти величины меняются со временем, поскольку любой детектор регистрирует лишь средние значения за промежутки времени, в течение которых возмущение очень много раз меняет знак. Наблюдаемая интенсивность $I(P)$ пропорциональна среднему значению $(V^{(r)}(P, t))^2$, и, следовательно, с точностью до несущественной постоянной получим, используя (10.2.20),

$$I(P) = 2 \langle (V^{(r)}(P, t))^2 \rangle = \langle V(P, t) V^*(P, t) \rangle. \quad (1)$$

Рассмотрим теперь две точки P_1 и P_2 в волновом поле. Можно экспериментально определить не только $I(P_1)$ и $I(P_2)$, но и интерференционные эффекты, возникающие при суперпозиции колебаний, исходящих из этих точек. Представим себе, что в исследуемое поле помещен непрозрачный экран $\mathcal{A}$ с небольшими отверстиями в P_1 и P_2, и рассмотрим распределение интенсивности на втором экране $\mathcal{B}$, находящемся на некотором расстоянии от $\mathcal{A}$ в направлении, противоположном направлению на источник (рис. 10.1). Для простоты будем считать, что показатель преломления среды между двумя экранами равен единице. Пусть s_1 и s_2 — расстояния от произвольной точки Q экрана $\mathcal{B}$ до P_1 и P_2. Точки

P_1 и P_2 можно считать центрами вторичных возмущений, так что комплексное возмущение в точке Q запишется в виде

$$V(Q,\ t)=K_1V(P_1,\ t-t_1)+K_2V(P_2,\ t-t_2). \tag{2}$$

Здесь t_1 и t_2 — времена распространения света от P_1 до Q и от P_2 до Q соответственно, т. е.

$$t_1=s_1/c,\ t_2=s_2/c, \tag{3}$$

где c — скорость света в вакууме. Коэффициенты K_1 и K_2 обратно пропорциональны s_1 и s_2 и, кроме того, зависят от размера отверстий и геометрии всего устройства (углов падения и углов дифракции в точках P_1 и P_2). Так как фаза вторичных волн, распространяющихся от P_1 и P_2, отличается от фазы первичной волны на четверть периода (см. §§ 8.2 и 8.3), K_1 и K_2 являются чисто мнимыми величинами.

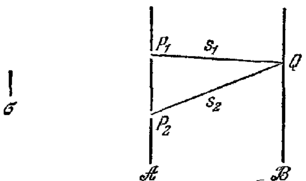


Рис. 10.1. Интерференционный эксперимент с полихроматическим светом от протяженного источника σ.

Из (1) и (2) следует *), что интенсивность в точке Q равна

$$I(Q)=K_1K_1^*\langle V_1(t-t_1)V_1^*(t-t_1)\rangle+K_2K_2^*\langle V_2(t-t_2)V_2^*(t-t_2)\rangle+$$
$$+K_1K_2^*\langle V_1(t-t_1)V_2^*(t-t_2)\rangle+K_2K_1^*\langle V_2(t-t_2)V_1^*(t-t_1)\rangle. \tag{4}$$

Поскольку предполагается, что поле стационарно **), во всех этих выражениях можно изменить начало отсчета времени. Тогда получим, например,

$$\langle V_1(t-t_1)V_1^*(t-t_1)\rangle=\langle V_1(t)V_1^*(t)\rangle=I_1. \tag{5}$$

Если использовать также (3) и вспомнить, что K_1 и K_2 — чисто мнимые величины, то (4) можно свести к

$$I(Q)=|K_1|^2I_1+|K_2|^2I_2+2|K_1K_2|\Gamma_{12}^{(r)}\left(\frac{s_2-s_1}{c}\right), \tag{6}$$

где $\Gamma_{12}^{(r)}(\tau)$ — вещественная часть функции

$$\Gamma_{12}(\tau)=\langle V_1(t+\tau)V_2^*(t)\rangle. \tag{7}$$

Понятие, определяемое выражением (7), служит основным в теории частичной когерентности. Мы будем называть его *взаимной когерентностью* световых колебаний в точках P_1 и P_2, причем колебания в точке P_1 рассматриваются в момент времени, запаздывающий на величину τ по сравнению с моментом времени колебаний в точке P_2. Мы будем называть функцию $\Gamma_{12}(\tau)$ *взаимной функцией когерентности* ***) волнового поля. Когда обе точки совпадают ($P_1=P_2$), получим

$$\Gamma_{11}(\tau)=\langle V_1(t+\tau)V_1^*(t)\rangle, \tag{8}$$

и тогда мы говорим об *автокогерентности* световых колебаний в P_1. При $\tau=0$ это соотношение сводится к выражению для обычной интенсивности

$$\Gamma_{11}(0)=I_1,\qquad \Gamma_{22}(0)=I_2.$$

---

*) В дальнейшем там, где это удобно, мы будем пользоваться более короткой записью, а именно $V_1(t)$ вместо $V(P_1,\ t)$, $\Gamma_{12}(\tau)$ вместо $\Gamma(P_1,\ P_2,\ \tau)$ и т. д.

**) Необходимое здесь предположение о стационарности равносильно утверждению, что $\langle V_i(t)\rangle$ не зависит от выбора начала отсчета времени и что корреляционные функции $\Gamma_{ij}(t',t'')=\langle V_i(t+t')V_j(t+t'')\rangle$ $(i,\ j=1,\ 2)$ зависят лишь от разности $t'-t''$. В этом случае мы часто говорим о *«стационарности в широком смысле»* (см. [47]).

***) В общей теории стационарных случайных процессов $\Gamma_{12}(\tau)$ называется *взаимной корреляционной функцией* величин $V_1(t)$ и $V_2(t)$, а $\Gamma_{11}(\tau)$ — *автокорреляционной функцией* величины $V_1(t)$.

Очевидно, что член $|K_1|^2 I_1$ в (6) представляет интенсивность, которая наблюдалась бы в Q, если бы открытым было лишь отверстие $P_1 (K_2=0)$; аналогичный смысл имеет и член $|K_2|^2 I_2$. Обозначим эти интенсивности соответственно через $I^{(1)}(Q)$ и $I^{(2)}(Q)$, т. е.

$$I^{(1)}(Q)=|K_1|^2 I_1=|K_1|^2\Gamma_{11}(0), \qquad I^{(2)}(Q)=|K_2|^2 I_2=|K_2|^2\Gamma_{22}(0), \tag{9}$$

а также нормируем $\Gamma_{12}(\tau)$, положив

$$\gamma_{12}(\tau)=\frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}}=\frac{\Gamma_{12}(\tau)}{\sqrt{I_1}\sqrt{I_2}}. \tag{10}$$

По причинам, которые вскоре станут ясными, назовем величину $\gamma_{12}(\tau)$ *комплексной степенью когерентности* световых колебаний. С помощью выражений (9) и (10) формулу (6) окончательно можно переписать в виде

$$I(Q)=I^{(1)}(Q)+I^{(2)}(Q)+2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,\gamma_{12}^{(r)}\left(\frac{s_2-s_1}{c}\right), \tag{11}$$

где $\gamma_{12}^{(r)}$— вещественная часть γ.

Последнее соотношение выражает *общий закон интерференции для стационарных оптических полей*. Оно показывает, что для определения интенсивности, возникающей при суперпозиции двух пучков света, необходимо знать интенсивность каждого пучка и значение вещественной части $\gamma_{12}^{(r)}$ комплексной степени когерентности. Позже мы покажем, как можно найти $\gamma_{12}^{(r)}$, зная характеристики источника и пропускание среды.

Выражение (11) остается справедливым при условии, что $s_2 - s_1$ заменено на разность хода $P_2Q - P_1Q$, если свет от P_1 и P_2 попадает в Q не прямо, а через промежуточную оптическую систему, и если можно пренебречь эффектами дисперсии. При таком обобщении формула (11) выполняется и в том случае, когда два интерферирующих пучка получаются из первичного пучка не путем «деления волнового фронта» в P_1 и P_2, а путем «деления амплитуды» в непосредственной близости к одной точке P_1, например в интерферометре Майкельсона. В последнем случае в соотношение (11) будет входить $\gamma_{11}^{(r)}(\tau)$ вместо $\gamma_{12}^{(r)}(\tau)$.

В отличие от возмущения $V^{(r)}$, корреляционные функции $\gamma_{12}^{(r)}$ и $\Gamma_{12}^{(r)}$ представляют величины, которые можно определить из эксперимента. Чтобы найти значение $\gamma_{12}^{(r)}$ для любой данной пары точек P_1 и P_2 и для любого заданного значения τ, помещают в световой пучок непрозрачный экран с небольшими отверстиями в P_1 и P_2, как показано на рис. 10.1, и измеряют интенсивность $I(Q)$ в некоторой точке Q позади экрана, для которой $P_2Q-P_1Q=c\tau$. Затем отдельно измеряют интенсивности $I^{(1)}(Q)$ и $I^{(2)}(Q)$ света, прошедшего через каждое отверстие. Выразив $\gamma_{12}^{(r)}$ через три найденные величины, получим, согласно (11)

$$\gamma_{12}^{(r)}=\frac{I(Q)-I^{(1)}(Q)-I^{(2)}(Q)}{2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}}. \tag{12}$$

Для определения $\Gamma_{12}^{(r)}$ следует также измерить интенсивности $I(P_1)$ и $I(P_2)$ в каждом отверстии. Тогда, согласно (10) и (12), получим для $\Gamma_{12}^{(r)}$

$$\Gamma_{12}^{(r)}=\sqrt{I(P_1)}\sqrt{I(P_2)}\,\gamma_{12}^{(r)}=\frac{1}{2}\sqrt{\frac{I(P_1)\,I(P_2)}{I^{(1)}(Q)\,I^{(2)}(Q)}}\left[I(Q)-I^{(1)}(Q)-I^{(2)}(Q)\right]. \tag{13}$$

Возвращаясь к (10), нетрудно увидеть, что при нашей нормировке $|\gamma_{12}(\tau)|\leqslant 1$. Чтобы показать это, введем, как в (10.2.17), обрезанные функции

$$\left.\begin{aligned} V_T^{(r)}(P,\ t)&=V^{(r)}(P,\ t), &&\text{когда}\quad |t|\leqslant T,\\ V_T^{(r)}(P,\ t)&=0, &&\text{когда}\quad |t|>T, \end{aligned}\right\} \tag{14}$$

и обозначим через $V_T(P,\ t)$ ассоциированный аналитический сигнал. Согласно

неравенству Шварца (см., например, [50], стр. 131) имеем

$$\left|\int_{-\infty}^{\infty} V_T(P_1,\ t+\tau)V_T^*(P_2,\ t)\,dt\right|^2 \leqslant$$
$$\leqslant \int_{-\infty}^{\infty} V_T(P_1,\ t+\tau)V_T^*(P_1,\ t+\tau)\,dt \int_{-\infty}^{\infty} V_T(P_2,\ t)V_T^*(P_2,\ t)\,dt. \quad (15)$$

В первом интеграле правой части $t+\tau$ можно заменить на t. Тогда, разделив обе части на $4T^2$ и переходя к пределу при $T\to\infty$, получим

$$|\Gamma_{12}(\tau)|^2 \leqslant \Gamma_{11}(0)\,\Gamma_{22}(0), \quad (16)$$

или, учитывая (10),

$$|\gamma_{12}(\tau)| \leqslant 1. \quad (17)$$

Смысл γ_{12} легче всего понять, если выразить (11) в несколько иной форме. Обозначим через $\bar{\nu}$ среднюю частоту света и запишем

$$\gamma_{12}(\tau) = |\gamma_{12}(\tau)|\exp\{i\,[\alpha_{12}(\tau)-2\pi\bar{\nu}\tau]\}, \quad (18)$$

где

$$\alpha_{12}(\tau) = 2\pi\bar{\nu}\tau + \arg\gamma_{12}(\tau). \quad (19)$$

Тогда (11) перейдет в

$$I(Q) = I^{(1)}(Q) + I^{(2)}(Q) + 2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\gamma_{12}(\tau)|\cos[\alpha_{12}(\tau)-\delta], \quad (20)$$

где параметр τ и разность фаз δ равны

$$\tau = \frac{s_2-s_1}{c}, \qquad \delta = 2\pi\bar{\nu}\tau = \frac{2\pi}{\bar{\lambda}}(s_2-s_1), \quad (21)$$

а $\bar{\lambda}$ — средняя длина волны. Если $|\gamma_{12}(\tau)|$ достигает своего максимального значения, равного единице, то интенсивность в точке Q будет совпадать с интенсивностью, которая получилась бы со строго монохроматическим светом с длиной волны $\bar{\lambda}$ и разностью фаз между колебаниями в P_1 и P_2, равной $\alpha_{12}(\tau)$. В таком случае можно сказать, что колебания в P_1 и P_2 (при соответствующем времени задержки τ между ними) *когерентны* *). Если $\gamma_{12}(\tau)$ имеет другое экстремальное значение, а именно нуль, последний член в (20) исчезает. Тогда не возникает никаких интерференционных эффектов от этих пучков, и можно сказать, что колебания *некогерентны*. Если $|\gamma_{12}(\tau)|$ не совпадает ни с одним из двух экстремальных значений, т. е. если $0<|\gamma_{12}(\tau)|<1$, то говорят, что колебания *частично когерентны*, причем $|\gamma_{12}(\tau)|$ представляет *степень их когерентности*.

При любом значении $|\gamma_{12}|$ интенсивность $I(Q)$ можно также представить в виде

$$I(Q) = |\gamma_{12}(\tau)|\,\{I^{(1)}(Q)+I^{(2)}(Q)+2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,\times$$
$$\times\cos[\alpha_{12}(\tau)-\delta]\} + \{1-|\gamma_{12}(\tau)|\}\,\{I^{(1)}(Q)+I^{(2)}(Q)\}. \quad (22)$$

Можно считать, что члены в первой строчке возникают из-за *когерентной* суперпозиции двух пучков с интенсивностями $|\gamma_{12}(\tau)|\,I^{(1)}(Q)$ и $|\gamma_{12}(\tau)|\,I^{(2)}(Q)$ и относительной разностью фаз $\alpha_{12}(\tau)-\delta$; члены во второй строчке — из-за *некогерентной* суперпозиции двух пучков с интенсивностями $[1-|\gamma_{12}(\tau)|]\,I^{(1)}(Q)$ и $[1-|\gamma_{12}(\tau)|]\,I^{(2)}(Q)$. Таким образом, свет от обоих отверстий, достигающий точки Q, можно считать как бы состоящим из смеси когерентной и некогерентной частей с отношением интенсивностей

$$\frac{I_{\text{ког}}}{I_{\text{неког}}} = \frac{|\gamma_{12}(\tau)|}{1-|\gamma_{12}(\tau)|}, \quad (23а)$$

*) Общие свойства когерентного света исследованы в работе [51].

или

$$\frac{I_{\text{ког}}}{I_{\text{поля}}} = |\gamma_{12}|, \qquad (I_{\text{поля}} = I_{\text{ког}} + I_{\text{неког}}). \tag{23б}$$

Мы видели (см. уравнение (12)), что $\gamma_{12}^{(r)}$ можно найти, измеряя интенсивности в соответствующем интерференционном эксперименте. Ниже (см. п. 10.4.1) мы покажем, что для большинства случаев, представляющих практический интерес, из таких экспериментов можно найти также и модуль (а в принципе и фазу) γ_{12}.

10.3.2. Спектральное представление взаимной когерентности. Пусть

$$V_T^{(r)}(P, t) = \int_{-\infty}^{\infty} v_T(P, \nu) \exp(-2\pi i \nu t)\, d\nu \tag{24}$$

— интегральное фурье-представление вещественной обрезанной функции $V_T^{(r)}$. Тогда обратное фурье-преобразование дает

$$v_T(P, \nu) = \int_{-\infty}^{\infty} V_T^{(r)}(P, t) \exp(2\pi i \nu t)\, dt, \tag{25}$$

откуда следует, что

$$\int_{-\infty}^{\infty} V_T^{(r)}(P_1, t+\tau) V_T^{(r)}(P_2, t)\, dt =$$
$$= \int_{-\infty}^{\infty} V_T^{(r)}(P_2, t) \left[\int_{-\infty}^{\infty} v_T(P_1, \nu) \exp[-2\pi i \nu (t+\tau)]\, d\nu \right] dt =$$
$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} V_T^{(r)}(P_2, t) \exp(-2\pi i \nu t)\, dt \right] v_T(P_1, \nu) \exp(-2\pi i \nu \tau)\, d\nu =$$
$$= \int_{-\infty}^{\infty} v_T(P_1, \nu)\, v_T^*(P_2, \nu) \exp(-2\pi i \nu \tau)\, d\nu. \tag{26}$$

Разделим обе части последнего соотношения на $2T$ и применим к величине $v_T(P_1, \nu)\, v_T^*(P_2, \nu)/2T$ «операцию сглаживания» по ансамблю случайных функций $V^{(r)}$. Такое усреднение по ансамблю (обозначавшееся чертой) было описано выше в связи с уравнением (10.2.20). Наконец, переходя к пределу $T \to \infty$, мы можем ожидать, что *)

$$\langle V^{(r)}(P_1, t+\tau) V^{(r)}(P_2, t) \rangle = \int_{-\infty}^{\infty} G_{12}(\nu) \exp(-2\pi i \nu \tau)\, d\nu, \tag{27}$$

где

$$G_{12}(\nu) = \lim_{T \to \infty} \left[\overline{\frac{v_T(P_1, \nu)\, v_T^*(P_2, \nu)}{2T}} \right]. \tag{28}$$

Функцию $G_{12}(\nu)$ можно назвать *взаимной спектральной плотностью* световых колебаний в точках P_1 и P_2. Она представляет собой обобщение *спектральной плотности*, введенной ранее (см. (10.2.22)), и переходит в нее при совпадении обеих точек. Понятие взаимной спектральной плотности является оптическим аналогом понятия *взаимного спектра мощности* в теории стационарных случайных процессов. Уравнение (27) показывает, что вещественная корреляционная функция $\langle V^{(r)}(P_1, t+\tau) V^{(r)}(P_2, t) \rangle$ и взаимная спектральная плотность $G_{12}(\nu)$ образуют пару, связанную фурье-преобразованием **).

*) Здесь справедливо то же примечание, что и на стр. 457.

**) Когда $P_1 = P_2$, этот результат является оптическим эквивалентом хорошо известной теоремы Винера — Хинчина [43, 52].

Вместо процедуры сглаживания для определения взаимной спектральной плотности $G_{12}(\nu)$ можно использовать обратное фурье-преобразование от (27). Для настоящей главы такой подход полностью эквивалентен процедуре, указанной в тексте.

Перейдем теперь к комплексному представлению. Пусть

$$V(P,\ t) = 2\int_0^\infty v(\nu)\exp(-2\pi i\nu t)\,d\nu \tag{29}$$

— аналитический сигнал (см. § 10.2), ассоциированный с $V^{(r)}(P,\ t)$. Пользуясь теми же приемами, при помощи которых мы перешли от (24) к (27), можно показать, что

$$\Gamma_{12}(\tau) = \langle V(P_1,\ t+\tau)V^*(P_2,\ t)\rangle = 4\int_0^\infty G_{12}(\nu)\exp(-2\pi i\nu\tau)\,d\nu. \tag{30}$$

Так как величина Γ_{12} не содержит спектральных компонент, принадлежащих отрицательным частотам, она представляет собой аналитический сигнал. Следовательно, если через $\Gamma_{12}^{(r)}$ и $\Gamma_{12}^{(i)}$ обозначить ее вещественную и мнимую части, т. е. считать, что

$$\Gamma_{12}(\tau) = \Gamma_{12}^{(r)}(\tau) + i\Gamma_{12}^{(i)}(\tau), \tag{31}$$

то эти функции будут связаны преобразованиями Гильберта, а именно

$$\Gamma_{12}^{(i)}(\tau) = \frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{\Gamma_{12}^{(r)}(\tau')}{\tau'-\tau}d\tau', \qquad \Gamma_{12}^{(r)}(\tau) = -\frac{1}{\pi}P\int_{-\infty}^{\infty}\frac{\Gamma_{12}^{(i)}(\tau')}{\tau'-\tau}d\tau'. \tag{32}$$

Отсюда вытекает, что величина $|\Gamma_{12}|$, рассматриваемая как функция τ, является огибающей $\Gamma_{12}^{(r)}$ (см. (10.2.10)—(10.2.15)), а из (30), (31) и (27) следует, что *)

$$\Gamma_{12}^{(r)}(\tau) = 2\langle V^{(r)}(P_1,\ t+\tau)V^{(r)}(P_2,\ t)\rangle = 2\int_{-\infty}^{\infty} G_{12}(\nu)\exp(-2\pi i\nu\tau)\,d\nu. \tag{33}$$

Кроме того, $|\gamma_{12}|$ служит огибающей вещественного коэффициента корреляции

$$\gamma_{12}^{(r)}(\tau) = \frac{\Gamma_{12}^{(r)}(\tau)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} = \frac{\langle V^{(r)}(P_1,\ t+\tau)V^{(r)}(P_2,\ t)\rangle}{\sqrt{\langle(V^{(r)}(P_1,\ t))^2\rangle}\sqrt{\langle(V^{(r)}(P_2,\ t))^2\rangle}}. \tag{34}$$

Уравнение (30) служит спектральным представлением взаимной функции когерентности $\Gamma_{12}(\tau)$. Уравнение (33) показывает, что вещественная часть $\Gamma_{12}(\tau)$ равна удвоенному значению взаимной корреляционной функции вещественных функций $V^{(r)}(P_1,\ t)$ и $V^{(r)}(P_2,\ t)$, а (32) определяет связь между вещественной и мнимой частями $\Gamma_{12}(\tau)$.

§ 10.4. Интерференция и дифракция квазимонохроматического света

Как мы видели, для адекватного описания интерференции частично когерентного света, вообще говоря, необходимо знать взаимную функцию когерентности $\Gamma_{12}(\tau)$ или, что эквивалентно этому, обычные интенсивности I_1 и I_2 и комплексную степень когерентности $\gamma_{12}(\tau)$. Здесь мы ограничимся важным случаем квазимонохроматического света, т. е. света, состоящего из спектральных компонент, которые занимают частотный интервал $\Delta\nu$, малый по сравнению со средней частотой $\bar\nu$. Мы покажем, что в этом случае теория принимает более простой вид. В частности, мы найдем, что при определенном дополнительном предположении, которое выполняется во многих приложениях, вместо $\Gamma_{12}(\tau)$ и $\gamma_{12}(\tau)$ можно применять корреляционные функции, не зависящие от параметра τ.

*) Нетрудно показать, что функция $\Gamma_{12}^{(r)}(\tau)$ равна также

$$2\langle V^{(i)}(P_1,\ t+\tau)V^{(i)}(P_2,\ t)\rangle$$

и

$$\Gamma_{12}^{(i)}(\tau) = 2\langle V^{(i)}(P_1,\ t+\tau)V^{(r)}(P_2,\ t)\rangle = -2\langle V^{(r)}(P_1,\ t+\tau)V^{(i)}(P_2,\ t)\rangle.$$

(См., например, [53] или [54], стр. 241—242).

10.4.1. Интерференция квазимонохроматического света. Взаимная интенсивность. Обратимся вновь к интерференционному эксперименту, изображенному на рис. 10.1. Согласно уравнению (10.3.20) интенсивность в точке Q интерференционной картины задается соотношением

$$I(Q)=I^{(1)}(Q)+I^{(2)}(Q)+2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\gamma_{12}(\tau)|\cos[\alpha_{12}(\tau)-\delta], \qquad (1)$$

где

$$\tau=\frac{s_2-s_1}{c}, \qquad \delta=2\pi\bar{\nu}\tau=\frac{2\pi}{\bar{\lambda}}(s_2-s_1). \qquad (2)$$

Допустим, что мы имеем здесь дело с квазимонохроматическим светом. Тогда из уравнения (10.3.18) следует (аналогично тому, как это было показано в связи с уравнением (10.2.11)), что по сравнению с $\cos 2\pi\bar{\nu}\tau$ и $\sin 2\pi\bar{\nu}\tau$ величины $|\gamma_{12}(\tau)|$ и $\alpha_{12}(\tau)$ будут медленно меняющимися функциями τ. Более того, если отверстия в P_1 и P_2 достаточно малы, интенсивности $I^{(1)}(Q)$ и $I^{(2)}(Q)$ света, дифрагировавшего на каждом отверстии по отдельности, сохраняются в достаточной степени постоянными в области, в которой $\cos 2\pi\bar{\nu}\tau$ и $\sin 2\pi\bar{\nu}\tau$ многократно меняют знак. Отсюда следует, что распределение интенсивности в окрестности любой точки Q слагается из почти однородного фона $I^{(1)}(Q)+I^{(2)}(Q)$

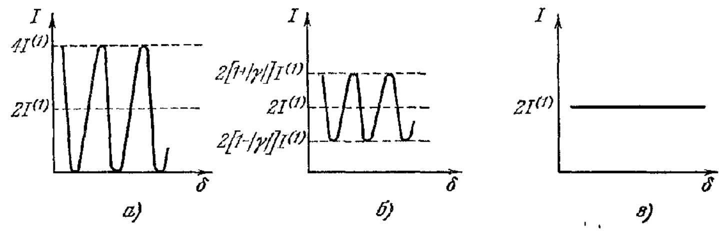


Рис. 10.2. Распределение интенсивности в интерференционной картине, образованной двумя квазимонохроматическими пучками равной интенсивности $I^{(1)}$ со степенью когерентности $|\gamma|$. *а* – когерентная суперпозиция ($|\gamma|=1$), *б* – частично когерентная суперпозиция ($0<|\gamma|<1$); *в* – некогерентная суперпозиция ($\gamma=0$).

и наложенного на него синусоидального распределения с почти постоянной амплитудой, равной $2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\gamma_{12}(\tau)|$. На рис. 10.2 показано распределение полной интенсивности для трех типичных случаев. Максимумы и минимумы интенсивности вблизи Q с хорошей точностью определяются выражениями

$$\left.\begin{aligned} I_{\text{макс}}&=I^{(1)}(Q)+I^{(2)}(Q)+2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\gamma_{12}(\tau)|,\\ I_{\text{мин}}&=I^{(1)}(Q)+I^{(2)}(Q)-2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\gamma_{12}(\tau)|. \end{aligned}\right\} \qquad (3)$$

Следовательно, *видность полос в точке Q равняется*

$$\mathcal{V}(Q)=\frac{I_{\text{макс}}-I_{\text{мин}}}{I_{\text{макс}}+I_{\text{мин}}}=\frac{2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}}{I^{(1)}(Q)+I^{(2)}(Q)}\,|\gamma_{12}(\tau)|. \qquad (4)$$

Эта формула связывает видность полос с интенсивностями двух пучков и их степенью когерентности. Если, как часто бывает, интенсивности обоих пучков равны $[I^{(1)}=I^{(2)}]$, то (4) переходит в

$$\mathcal{V}(Q)=|\gamma_{12}(\tau)|, \qquad (5)$$

т. е. в этом случае видность полос равна степени когерентности.

Согласно (1) и (2) положения максимумов интенсивности вблизи Q определяются выражением

$$\frac{2\pi}{\bar{\lambda}}(s_2-s_1)-\alpha_{12}(\tau)=2m\pi \qquad (m=0,\ \pm 1,\ \pm 2,\ \ldots).$$

Таким же выражением определялись бы положения максимумов при освещении отверстий строго монохроматическим светом с длиной волны $\bar{\lambda}$, если бы фаза колебаний в P_1 отставала относительно фазы в P_2 на $\alpha_{12}(\tau)$. Согласно (7.3.7) отставание фазы на величину 2π соответствует смещению интерференционной картины в направлении, параллельном P_1P_2, на величину $a\bar{\lambda}/d$, где d — расстояние между P_1 и P_2, а a — расстояние между экранами $\mathcal{A}$ и $\mathcal{B}$. Следовательно, *полосы, полученные с квазимонохроматическим светом, смещены относительно полос, которые образовались бы при синфазном освещении точек P_1 и P_2 монохроматическим светом на величину*

$$x = \frac{\bar{\lambda}}{2\pi}\frac{a}{d}\alpha_{12}(\tau) \tag{6}$$

в направлении, параллельном линии, соединяющей отверстия.

Мы видим, что из измерений видности и положения интерференционных полос можно определить амплитуду и фазу комплексной степени когерентности квазимонохроматических пучков света. Метод их определения тесно связан с описанным в п. 7.5.8 методом Майкельсона для определения распределения интенсивности в спектральных линиях из измерений кривых видности. Из уравнений (10.3.10) и (10.3.30) следует, что

$$\gamma_{11}(\tau) = \frac{\int\limits_0^\infty G(\nu)\exp(-2\pi i\nu\tau)\,d\nu}{\int\limits_0^\infty G(\nu)\,d\nu},$$

где $G(\nu)$ — спектральная плотность. Следовательно, согласно обратной теореме Фурье, $G(\nu)$ пропорциональна фурье-образу величины $\gamma_{11}(\tau)$. Но мы только что видели, что модуль $\gamma_{11}(\tau)$ по существу равен видности полос, образующихся в соответствующем интерференционном эксперименте, а фаза $\gamma_{11}(\tau)$ связана с их положением простым соотношением. Таким образом, мы приходим к такому же способу расчета G, каким пользовался Майкельсон. Очевидно, можно считать, что кривые видности, приведенные на рис. 7.54 и 7.55, представляют $|\gamma_{11}|$ как функцию разности хода двух интерферирующих пучков.

На практике время задержки τ одного интерферирующего пучка относительно другого часто довольно мало, и тогда приведенные выше формулы легко упростить. Согласно уравнениям (10.3.10), (10.3.18) и (10.3.30) имеем

$$|\Gamma_{12}(\tau)|\exp[i\alpha_{12}(\tau)] = \sqrt{I_1}\sqrt{I_2}\,|\gamma_{12}(\tau)|\exp[i\alpha_{12}(\tau)] =$$
$$= 4\int\limits_0^\infty G_{12}(\nu)\exp[-2\pi i(\nu-\bar{\nu})\tau]\,d\nu. \tag{7}$$

Если модуль $|\tau|$ так мал, что $|(\nu-\bar{\nu})\tau|\ll 1$ для всех частот, при которых $|G_{12}(\nu)|$ имеет заметную величину, т. е. если

$$|\tau| \ll \frac{1}{\Delta\nu}, \tag{8}$$

то, очевидно, мы внесем лишь небольшую ошибку, если заменим экспоненциальный член подынтегрального выражения в (7) единицей. Условие (8) означает, что, согласно (7.5.105), $|\tau|$ должен быть мал по сравнению с временем когерентности света. При этом условии $|\Gamma_{12}(\tau)|$, $|\gamma_{12}(\tau)|$ и $\alpha_{12}(\tau)$ незначительно

отличаются от $|\Gamma_{12}(0)|$, $|\gamma_{12}(0)|$ и $\alpha_{12}(0)$ соответственно. Удобно положить *)

$$J_{12}=\Gamma_{12}(0)=\langle V_1(t)V_2^*(t)\rangle, \tag{9а}$$

$$\mu_{12}=\gamma_{12}(0)=\frac{\Gamma_{12}(0)}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}}=\frac{J_{12}}{\sqrt{J_{11}}\sqrt{J_{22}}}=\frac{J_{12}}{\sqrt{I_1}\sqrt{I_2}}, \tag{9б}$$

$$\beta_{12}=\alpha_{12}(0)=\arg\gamma_{12}(0)=\arg\mu_{12}. \tag{9в}$$

Теперь уравнения (10.3.10) и (10.3.18) дают при выполнении условия (8)

$$\gamma_{12}(\tau)\approx|\mu_{12}|\exp[i(\beta_{12}-2\pi\bar\nu\tau)]=\mu_{12}\exp(-2\pi i\bar\nu\tau), \tag{10а}$$

$$\Gamma_{12}(\tau)\approx|J_{12}|\exp[i(\beta_{12}-2\pi\bar\nu\tau)]=J_{12}\exp(-2\pi i\bar\nu\tau). \tag{10б}$$

Таким образом, *при выполнении условия* (8) *во всех наших формулах мы можем заменить* $\gamma_{12}(\tau)$ *и* $\Gamma_{12}(\tau)$ *на величины, стоящие в правых частях соотношений* (10а) *и* (10б) *соответственно. В частности, закон интерференции* (1) *примет вид*

$$I(Q)\approx I^{(1)}(Q)+I^{(2)}(Q)+2\sqrt{I^{(1)}(Q)}\sqrt{I^{(2)}(Q)}\,|\mu_{12}|\cos(\beta_{12}-\delta). \tag{11}$$

Он будет выполняться до тех пор, пока разность хода $|s_2-s_1|=c|\tau|$ между интерферирующими пучками будет мала по сравнению с длиной когерентности $c/\Delta\nu$, т. е. до тех пор, пока

$$|\Delta\mathscr{S}|=|s_2-s_1|=\frac{\bar\lambda}{2\pi}\delta\ll\frac{\bar\lambda^2}{\Delta\lambda}, \tag{12}$$

где использовано соотношение $c/\Delta\nu=\bar\lambda^2/\Delta\lambda$.

Уравнение (11) является основной формулой элементарной (квазимонохроматической) теории частичной когерентности. Эта теория составит предмет рассмотрения в оставшейся части настоящего параграфа; в § 10.5 будут рассмотрены некоторые ее приложения. Если справедливо уравнение (11) (т. е. выполнены неравенства (8) или (12)), то корреляция между колебаниями в любых двух точках P_1 и P_2 волнового поля характеризуется не $\Gamma_{12}(\tau)$, а J_{12}, т. е. величиной, которая зависит не от разности времен τ, а от положения этих точек. В пределах применимости элементарной теории мы можем написать, как видно из (10а),

$$|\gamma_{12}(\tau)|\approx|\mu_{12}|, \tag{13}$$

так что $|\mu_{12}|$ $(0\leqslant|\mu_{12}|\leqslant1)$ представляет степень когерентности колебаний в точках P_1 и P_2. Из уравнения (11) следует, что фаза β_{12} величины μ_{12} представляет собой эффективную разность фаз этих колебаний. Величину μ_{12} (так же, как и $\gamma_{12}(\tau)$, частным случаем которой она является) обычно называют *комплексной степенью когерентности* (иногда *комплексным коэффициентом когерентности*), *величину* J_{12} — *взаимной интенсивностью*.

10.4.2. Расчет взаимной интенсивности и степени когерентности для света от протяженного некогерентного квазимонохроматического источника. *а. Теорема Ван-Циттерта — Цернике.* Определим взаимную интенсивность J_{12} и комплексную степень когерентности μ_{12} для точек P_1 и P_2 экрана $\mathscr{A}$, освещаемого протяженным квазимонохроматическим первичным источником σ. Для простоты в качестве σ возьмем часть плоскости, параллельной $\mathscr{A}$, и предположим, что среда между источником и экраном однородна. Допустим также, что малы как линейные размеры σ по сравнению с расстоянием OO' между источником и экраном (рис. 10.3), так и углы между OO' и линиями, соединяющими произвольную точку S источника с точками P_1 и P_2.

Вообразим, что источник разделен на элементы $d\sigma_1$, $d\sigma_2$, ... с линейными размерами, малыми по сравнению со средней длиной волны $\bar\lambda$, и центрами, на-

*) Мы вновь пользуемся сокращенным обозначением, т. е. пишем J_{12} вместо $J(P_1, P_2)$ и т. д.

ходящимися в точках S_1, S_2, ... Если $V_{m1}(t)$ и $V_{m2}(t)$ — комплексные возмущения в точках P_1 и P_2, обусловленные элементом $d\sigma_m$, то общее возмущение в этих точках равно

$$V_1(t)=\sum_m V_{m1}(t), \qquad V_2(t)=\sum_m V_{m2}(t). \tag{14}$$

Следовательно,

$$J(P_1,\ P_2)=\langle V_1(t)V_2^*(t)\rangle=\sum_m\langle V_{m1}(t)V_{m2}^*(t)\rangle+\sum\sum_{m\neq n}\langle V_{m1}(t)V_{n2}^*(t)\rangle. \tag{15}$$

Рис. 10.3. К теореме Ван-Циттерта—Цернике.

Световые колебания, создаваемые различными элементами источника, можно считать статистически независимыми (взаимно некогерентными), причем среднее значение поля равно нулю; тогда *)

$$\langle V_{m1}(t)V_{n2}^*(t)\rangle=\langle V_{m1}(t)\rangle\langle V_{n2}^*(t)\rangle=0, \quad \text{когда } m\neq n. \tag{16}$$

Если R_{m1} и R_{m2} — расстояния точек P_1 и P_2 до элемента источника $d\sigma_m$, то

$$\left.\begin{aligned} V_{m1}(t)&=A_m\left(t-\frac{R_{m1}}{v}\right)\frac{\exp\left[-2\pi i\bar{\nu}\,(t-R_{m1}/v)\right]}{R_{m1}},\\ V_{m2}(t)&=A_m\left(t-\frac{R_{m2}}{v}\right)\frac{\exp\left[-2\pi i\bar{\nu}\,(t-R_{m2}/v)\right]}{R_{m2}},\end{aligned}\right\} \tag{17}$$

где $|A_m|$ характеризует силу, а $\arg A_m$ — фазу излучения от m-го элемента**), а v — скорость света в среде между источником и экраном. Следовательно,

$$\begin{aligned}\langle V_{m1}(t)\,V_{m2}^*(t)\rangle&=\left\langle A_m\left(t-\frac{R_{m1}}{v}\right)A_m^*\left(t-\frac{R_{m2}}{v}\right)\right\rangle\frac{\exp\left\{\dfrac{2\pi i\bar{\nu}\,(R_{m1}-R_{m2})}{v}\right\}}{R_{m1}R_{m2}}=\\ &=\left\langle A_m(t)\,A_m^*\left(t-\frac{R_{m2}-R_{m1}}{v}\right)\right\rangle\frac{\exp\left\{\dfrac{2\pi i\bar{\nu}\,(R_{m1}-R_{m2})}{v}\right\}}{R_{m1}R_{m2}}.\end{aligned} \tag{18}$$

Если разность хода $R_{m2}-R_{m1}$ мала по сравнению с длиной когерентности света, в аргументе A_m^* можно пренебречь запаздыванием $(R_{m2}-R_{m1})/v$. Тогда из (15), (16) и (18) получим

$$J(P_1,\ P_2)=\sum_m\langle A_m(t)\,A_m^*(t)\rangle\frac{\exp\left\{\dfrac{2\pi i\bar{\nu}\,(R_{m1}-R_{m2})}{v}\right\}}{R_{m1}R_{m2}}. \tag{19}$$

Величина $\langle A_m(t)\,A_m^*(t)\rangle$ характеризует интенсивность излучения, испускаемого элементом источника $d\sigma_m$. В любом практически интересном случае число элементов источника можно считать настолько большим, что мы вправе рассматривать источник как непрерывный. Обозначая через $I(S)$ интенсивность на единицу площади источника, т. е. $I(S_m)\,d\sigma_m=\langle A_m(t)\,A_m^*(t)\rangle$, получим ***)

*) О некогерентности можно говорить лишь при наличии конечного (хотя не обязательно широкого) спектрального интервала; поэтому уравнение (16) неверно для идеализированного случая строго монохроматического света. Для монохроматического света $V_{m1}(t)=U_{m1}\exp(-2\pi i\nu t)$, $V_{n2}(t)=U_{n2}\exp(-2\pi i\nu t)$, где U_{m1} и U_{n2} не зависят от времени, так что $\langle V_{m1}(t)V_{n2}^*(t)\rangle=U_{m1}U_{n2}^*$, а эта величина в общем случае отлична от нуля.

**) В общем случае A_m зависит также от направления, но для простоты мы пренебрежем этой зависимостью.

***) В дальнейшем мы будем часто пользоваться обозначениями dS, dP_1, ... для элементов поверхности с центрами в точках S, P_1, ...

вместо (19)

$$J(P_1P_2)=\int_\sigma I(S)\frac{\exp\{i\bar{k}(R_1-R_2)\}}{R_1R_2}dS, \tag{20}$$

где R_1 и R_2 — расстояния между произвольной точкой S источника и точками P_1 и P_2, а $\bar{k}=2\pi\bar{\nu}/v=2\pi/\bar{\lambda}$ — волновое число в среде. Комплексная степень когерентности $\mu(P_1, P_2)$, согласно (20) и (9б), равна

$$\mu(P_1, P_2)=\frac{1}{\sqrt{I(P_1)}\sqrt{I(P_2)}}\int_\sigma I(S)\frac{\exp\{i\bar{k}(R_1-R_2)\}}{R_1R_2}dS, \tag{21}$$

где

$$I(P_1)=J(P_1, P_1)=\int_\sigma\frac{I(S)}{R_1^2}dS, \quad I(P_2)=J(P_2, P_2)=\int_\sigma\frac{I(S)}{R_2^2}dS \tag{21a}$$

— интенсивности в точках P_1 и P_2.

Мы видим, что интеграл (21) совпадает с интегралом, который появляется в другом случае, а именно при вычислении на основе принципа Гюйгенса — Френеля комплексного возмущения в дифракционной картине, возникающей при дифракции сферической волны на отверстии в непрозрачном экране. Точнее, (21) означает, что *комплексная степень когерентности, которая описывает корреляцию колебаний в фиксированной точке P_2 и переменной точке P_1 плоскости, освещенной протяженным квазимонохроматическим первичным источником, равна нормированной комплексной амплитуде в соответствующей точке P_1 некоторой дифракционной картины с центром в точке P_2. Эта картина получится, если заменить источник дифракционным отверстием такого же размера и формы и заполнить его сферической волной, сходящейся в P_2, причем распределение амплитуд по волновому фронту в отверстии должно быть пропорциональным распределению интенсивности по источнику.* Этот результат впервые был получен Ван-Циттертом [8], а позднее, более простым способом, Цернике [11]. Мы будем именовать его *теоремой Ван-Циттерта — Цернике.*

В большинстве приложений можно считать, что интенсивность $I(S)$ не зависит от положения точки S на поверхности (постоянная интенсивность). Тогда соответствующая дифракционная проблема совпадает с проблемой дифракции сферической волны постоянной амплитуды на отверстии такого же размера и формы, как и источник.

Пусть (ξ, η) — координаты произвольной точки S источника в системе с началом в точке O, и пусть (X_1, Y_1) и (X_2, Y_2) — координаты точек P_1 и P_2 в системе с началом в точке O' и осями, параллельными осям первой системы (см. рис. 10.3). Тогда, если R — расстояние OO', то

$$R_1^2=(X_1-\xi)^2+(Y_1-\eta)^2+R^2,$$

так что

$$R_1\approx R+\frac{(X_1-\xi)^2+(Y_1-\eta)^2}{2R}. \tag{22}$$

Здесь оставлены лишь основные члены относительно X_1/R, Y_1/R, ξ/R и η/R. Для R_2 получается точно такое же выражение и, следовательно,

$$R_1-R_2\approx\frac{(X_1^2+Y_1^2)-(X_2^2+Y_2^2)}{2R}-\frac{(X_1-X_2)\xi+(Y_1-Y_2)\eta}{R}. \tag{23}$$

В знаменателе подынтегральных выражений (20) и (21) R_1 и R_2 с достаточно хорошей точностью можно заменить на R. Положим также

$$\frac{X_1-X_2}{R}=p, \quad \frac{Y_1-Y_2}{R}=q, \tag{24}$$

$$\psi=\frac{\bar{k}[(X_1^2+Y_1^2)-(X_2^2+Y_2^2)]}{2R}. \tag{25}$$

Тогда (21) примет вид

$$\mu_{12}=\frac{e^{i\psi}\iint\limits_{\sigma} I(\xi,\eta)\exp\left[-i\bar{k}(p\xi+q\eta)\right]d\xi\,d\eta}{\iint\limits_{\sigma} I(\xi,\eta)\,d\xi\,d\eta}. \qquad (26)$$

Следовательно, *если линейные размеры источника и расстояние между P_1 и P_2 малы по сравнению с расстоянием этих точек от источника, степень когерентности* $|\mu_{12}|$ *равна абсолютному значению нормированного преобразования Фурье от функции, описывающей интенсивность источника.*

Величина ψ, определяемая (25), допускает простую интерпретацию. Согласно (23) она представляет собой разность фаз $2\pi(OP_1-OP_2)/\bar{\lambda}$, которой, очевидно, можно пренебречь, когда

$$OP_1-OP_2\ll\bar{\lambda}. \qquad (27)$$

Для однородного источника в виде круга радиуса ρ с центром в точке O, интегрируя (26), получим (см. п. 8.5.2)

$$\mu_{12}=\left(\frac{2J_1(v)}{v}\right)e^{i\psi}, \qquad (28)$$

где

$$\left.\begin{aligned} v&=\bar{k}\rho\sqrt{p^2+q^2}=\frac{2\pi}{\bar{\lambda}}\frac{\rho}{R}\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2},\\ \psi&=\frac{2\pi}{\bar{\lambda}}\left[\frac{(X_1^2+Y_1^2)-(X_2^2+Y_2^2)}{2R}\right],\end{aligned}\right\} \qquad (29)$$

а J_1 — функция Бесселя первого рода первого порядка *). Согласно п. 8.5.2 величина $|2J_1(v)/v|$ монотонно уменьшается от значения, равного единице при $v=0$, до нулевого значения при $v=3{,}83$. Таким образом, степень когерентности монотонно уменьшается, когда точки P_1 и P_2 удаляются друг от друга. При удалении P_1 и P_2 на расстояние

$$P_1P_2=\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}=\frac{0{,}61R\bar{\lambda}}{\rho} \qquad (30)$$

достигается полная некогерентность. При дальнейшем увеличении v вновь возникает небольшая когерентность, но степень ее остается меньше 0,14. Полная некогерентность вновь наступает при $v=7{,}02$. Проходя через нуль, функция $J_1(v)$ каждый раз меняет знак, т. е. фаза $\beta_{12}=\arg\mu_{12}$ при этом изменяется на π. Следовательно, после каждого исчезновения полос яркие и темные полосы меняются местами.

Функция $|2J_1(v)/v|$ монотонно уменьшается от значения, равного единице при $v=0$, до 0,88 при $v=1$, т. е. при

$$P_1P_2=\frac{0{,}16R\bar{\lambda}}{\rho}. \qquad (31)$$

Считая отклонение в 12% максимально допустимым отклонением от идеального значения, равного единице, получим, что *квазимонохроматический однородный источник углового радиуса* $\alpha=\rho/R$ *почти когерентно освещает площадку в виде круга диаметром* **) $0{,}16\,\bar{\lambda}/\alpha$. Этот результат полезен при оценке

*) Никакой путаницы не должно возникать оттого, что символ J используется также для обозначения взаимной интенсивности, так как последняя встречается всегда с двумя индексами или с несколькими аргументами.

**) Верде еще в 1865 г. нашел, что диаметр «кружка когерентности» несколько меньше $0{,}5\,R\bar{\lambda}/\rho$ [1].

размеров источника, требуемых при проведении экспериментов по интерференции и дифракции.

В качестве примера рассмотрим размеры «области когерентности» вокруг произвольной точки экрана, освещаемого непосредственно Солнцем. Угол 2α, под которым солнечный диск виден на поверхности Земли, составляет около $0°32' \approx 0{,}018$ *рад*. Следовательно, если пренебречь изменением яркости по поверхности солнечного диска, диаметр d области когерентности приблизительно равен $0{,}16\,\bar{\lambda}/0{,}009 \approx 18\bar{\lambda}$. Для средней длины волны $\bar{\lambda} = 5{,}5 \cdot 10^{-5}$ *см* получим $d \approx 0{,}01$ *мм*.

В связи с изложенным выше предстает в новом свете метод Майкельсона измерения угловых диаметров звезд (см. п. 7.3.6). Согласно (5) и (13) видность полос равна степени когерентности световых колебаний на двух внешних зеркалах (M_1 и M_2 на рис. 7.16) звездного интерферометра Майкельсона. Для звездного диска в виде круга постоянной яркости с угловым радиусом α наименьшее разделение зеркал, при котором степень когерентности обращается в нуль (первое исчезновение полос), равно, согласно (30), $0{,}61\bar{\lambda}/\alpha$, что соответствует (7.3.42). Более того, из измерений видности и положения полос в принципе можно определить не только диаметр звезды, но и распределение интенсивности по ее диску. В самом деле, согласно п. 10.4.1, измерения видности и положения полос эквивалентны определению как амплитуды, так и фазы комплексной степени когерентности μ_{12}, а согласно (26) распределение интенсивности пропорционально обратному фурье-преобразованию μ_{12}.

В п. 7.3.6 мы упоминали о важной модификации звездного интерферометра Майкельсона, предложенной Брауном и Твиссом. В разработанной ими системе свет от звезды фокусируют на два фотоэлектрических детектора P_1 и P_2, и информация о звезде получается путем изучения корреляции флуктуаций их выходных токов. Полный анализ характеристик такой системы должен учитывать квантовую природу фотоэффекта *); он требует также определенных знаний по электронике и поэтому выходит за рамки настоящей книги. Однако нетрудно понять принцип метода. При идеальных условиях эксперимента (отсутствие шума) ток на выходе каждого фотоэлектрического детектора пропорционален мгновенной интенсивности $I(t)$ падающего света, а флуктуация этого тока пропорциональна $\Delta I(t) = I(t) - \langle I(t) \rangle$. Следовательно, в интерферометре Брауна и Твисса измеряется величина, пропорциональная $\Omega_{12} = \langle \Delta I_1 \Delta I_2 \rangle$. Простой статистический расчет показывает [59] (см. также [60]), что Ω_{12} пропорционально квадрату степени когерентности и, значит, величина Ω_{12}, так же как и $|\mu_{12}|$, дает информацию о размере звезды.

б. *Формула Гопкинса.* При выводе формулы Ван-Циттерта — Цернике (21) мы считали, что среда между источником σ и точками P_1 и P_2 однородна. Нетрудно обобщить эту формулу на другие случаи, например на случай неоднородной среды или среды, состоящей из ряда однородных областей с разными показателями преломления.

Вновь представим себе, что источник разделен на небольшие элементы $d\sigma_1$, $d\sigma_2$, ... с центрами в точках S_1, S_2, ..., причем линейные размеры этих элементов малы по сравнению со средней длиной волны $\bar{\lambda}$. Если, как и раньше, $V_{m1}(t)$ и $V_{m2}(t)$ — возмущения в точках P_1 и P_2, обусловленные элементом $d\sigma_m$, то уравнения (15) и (16) остаются по-прежнему справедливыми, но в (17) каждый множитель $(\exp i\bar{k}R_{mj})/R_{mj}$ ($j = 1, 2$; $\bar{k} = 2\pi\bar{\nu}/v$) необходимо заменить более общей функцией. Введем *функцию пропускания* среды $K(S, P, \nu)$, определяемую так же, как это было сделано в п. 9.5.1. Она представляет комплексное возмущение в точке P, обусловленное монохроматическим точечным источником с единичной силой и нулевой фазой, испускающим излучение частоты ν и расположенным в точке S элемента $d\sigma$. Для однородной среды из принципа

*) См. [55]. См. также [54, 56—58].

Гюйгенса — Френеля находим, что $K(S,\ P,\ \nu) = -(i \exp i\bar{k}R)/\bar{\lambda}R$, где R — расстояние SP; при этом предполагается, что угол между SP и нормалью к $d\sigma$ достаточно мал. В более общем случае множитель $\exp i\bar{k}R_{mj}/R_{mj}$ необходимо заменить на $i\bar{\lambda}K(S_m,\ P_j,\ \bar{\nu})$, и тогда, переходя к непрерывному распределению, вместо (20) мы получим следующее соотношение:

$$J(P_1,\ P_2) = \bar{\lambda}^2 \int_\sigma I(S)\, K(S,\ P_1,\ \bar{\nu})\, K^*(S,\ P_2,\ \bar{\nu})\, dS. \qquad (32)$$

Согласно (32) и (9б) имеем

$$\mu(P_1,\ P_2) = \frac{\bar{\lambda}^2}{\sqrt{I(P_1)}\sqrt{I(P_2)}} \int_\sigma I(S)\, K(S,\ P_1,\ \bar{\nu})\, K^*(S,\ P_2,\ \bar{\nu})\, dS, \qquad (33)$$

где $I(P_1) = J(P_1, P_1)$ и $I(P_2) = J(P_2, P_2)$ — интенсивности в точках P_1 и P_2 соответственно.

Для дальнейшего полезно представить последние два соотношения в несколько иной форме. Положим

$$i\bar{\lambda}K(S,\ P_1,\ \bar{\nu})\sqrt{I(S)} = U(S,\ P_1), \quad i\bar{\lambda}K(S,\ P_2,\ \bar{\nu})\sqrt{I(S)} = U(S,\ P_2). \qquad (34)$$

Тогда формулы (32) и (33) примут вид

$$J(P_1,\ P_2) = \int_\sigma U(S,\ P_1)\, U^*(S,\ P_2)\, dS, \qquad (35а)$$

$$\mu(P_1,\ P_2) = \frac{1}{\sqrt{I(P_1)}\sqrt{I(P_2)}} \int_\sigma U(S,\ P_1)\, U^*(S,\ P_2)\, dS. \qquad (35б)$$

Отметим, что определяемая соотношением (34) величина $U(S,\ P)$ пропорциональна возмущению, которое возникало бы в точке P от строго монохроматического точечного источника, испускающего излучение частотой $\bar{\nu}$ (с силой $\sqrt{I(S)}$ и нулевой фазой) и расположенного в точке S. Таким образом, можно рассматривать формулы (35) как соотношения, выражающие взаимную интенсивность $J(P_1,\ P_2)$ и комплексную степень когерентности $\mu(P_1,\ P_2)$, обусловленные протяженным квазимонохроматическим источником, через *возмущения*, создаваемые в P_1 и P_2 каждой точкой ассоциированного монохроматического источника *).

Выражение (35б) было впервые предложено Гопкинсом [12], исходившим из эвристических предположений. Оно чрезвычайно полезно при решении проблем когерентности в инструментальной оптике. Основная ценность этой формулы состоит в том, что так же, как и теорема Ван-Циттерта — Цернике, она позволяет рассчитать комплексную степень когерентности света, испускаемого некогерентным источником, без явного использования процесса усреднения.

10.4.3. Пример. В качестве иллюстрации к предыдущим рассуждениям рассмотрим такой эксперимент. Первичный источник σ_0 с помощью линзы L_0 изображается на небольшое отверстие σ_1. Пучок света, выходящий из этого отверстия, превращается в параллельный линзой L_1. Вторая линза L_2, в точности подобная L_1, сводит интерферирующие пучки в фокусе F своей фокальной плоскости $\mathcal{F}$. Плоское зеркало M служит для уменьшения общей длины прибора (рис. 10.4.).

*) Нельзя предположить, что для такого воображаемого источника *взаимная интенсивность и комплексная степень когерентности* света также будут определяться формулами (35). Как отмечалось выше, соотношение (16), использованное при выводе этих формул, несправедливо в предельном случае монохроматического излучения; степень когерентности монохроматического света фактически всегда равна единице.

Если в параллельном пучке между линзами L_1 и L_2 расположить дифракционную маску (темный экран $\mathcal{A}$, например, кусок равномерно почерневшей пленки) с отверстиями любого желаемого размера и формы и с любым их распределением, то в фокальной плоскости $\mathcal{F}$ получится ее фраунгоферовская дифракционная картина. Обычно оператор может смещать маску, рассматривая в то же время дифракционную картину в микроскоп *).

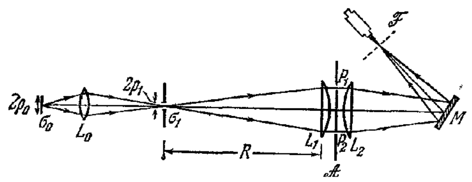


Рис. 10.4. Схема дифрактометра.

Если в качестве σ_0 использовать квазимонохроматический точечный источник, то вблизи его геометрического изображения в плоскости σ_1 получилось бы когерентное освещение. Размер такой когерентно освещаемой области порядка эффективного размера дифракционной картины Эйри σ_A, образуемой линзой L_0 и создаваемой одной точкой источника. Можно считать, что распределение света в плоскости σ_1, обусловленное конечным первичным источником, возникло в результате некогерентной суперпозиции ряда таких картин. Если, мы предположим, что изображение этого протяженного источника, образуемое L_0, и отверстие σ_1 велики по сравнению с σ_A, то освещенное отверстие σ_1 само служит *некогерентным источником* **). Согласно теореме Ван-Циттерта — Цернике при таком источнике будет существовать корреляция между колебаниями в любых двух точках на первой поверхности линзы L_1 (или, для большей общности, в плоскости $\mathcal{A}$). При обычных приближениях находим, что комплексная степень когерентности определяется формулой (28), т. е.

$$\mu_{12} = |\mu_{12}| \exp(i\beta_{12}) = \frac{2J_1(v)}{v} \exp(i\psi), \tag{36а}$$

где

$$v = \frac{2\pi}{\lambda} \frac{\rho_1 d}{R}, \qquad \psi = \frac{2\pi}{\lambda} \left(\frac{r_1^2 - r_2^2}{2R} \right), \tag{36б}$$

$d = P_1P_2$, ρ_1 — радиус σ_1, R — расстояние между σ_1 и L_1, r_1 и r_2 — расстояния от P_1 и P_2 до оси.

Если дифракционная маска $\mathcal{A}$ содержит два небольших круглых отверстия с центрами в точках P_1 и P_2, то получающаяся картина, наблюдаемая в фокальной плоскости $\mathcal{F}$, обусловлена суперпозицией двух выходящих из этих отверстий частично когерентных пучков со степенью когерентности $|\mu_{12}|$. Рассмотрим изменения в структуре этой картины при постепенном увеличении расстояния между P_1 и P_2, т. е. при изменении степени когерентности между двумя интерферирующими пучками.

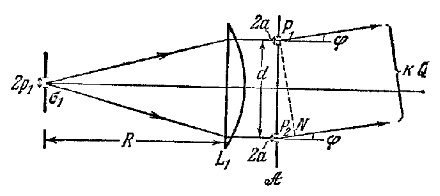


Рис. 10.5. К расчету распределения интенсивности в фокальной плоскости дифрактометра.

Предположим, что точки P_1 и P_2 расположены симметрично относительно оси. Тогда $\psi = 0$, а интенсивности $I^{(1)}(Q)$ и $I^{(2)}(Q)$ в точке Q фокальной плоскости, связанные с каждым из двух пучков, равны и выражаются с помощью формулы Фраунгофера для дифракции на круглом отверстии (8.5.14). Если точка Q

*) Этот прибор известен как *дифрактометр* и используется главным образом в оптических дифракционных методах для решения проблем рентгеноструктурного анализа (см. [61—63])

**) Этот вопрос рассматривается количественно в п. 10.5.1 (см. также [64]).

служит фокусом для лучей, дифрагировавших в направлениях, которые образуют угол φ с нормалью к $\mathcal{A}$, и если a — радиус каждого отверстия (рис. 10.5), то с точностью до нормирующего множителя имеем

$$I^{(1)}(Q)=I^{(2)}(Q)=\left(\frac{2J_1(u)}{u}\right)^2, \qquad u=\frac{2\pi}{\bar{\lambda}}a\sin\varphi. \tag{37}$$

Разность фаз δ для пучков, дифрагировавших к Q, равна

$$\delta=\frac{2\pi}{\bar{\lambda}}P_2N=\frac{2\pi}{\bar{\lambda}}d\sin\varphi=Cuv, \qquad C=\frac{\bar{\lambda}}{2\pi}\frac{R}{\rho_1 a}, \tag{38}$$

где N — основание перпендикуляра, опущенного из P_1 на луч, идущий из точки P_2. Подставляя (36), (37) и (38) в (11), окончательно получим для интенсивности в точке $Q(\varphi)$ фокальной плоскости, когда отверстия P_1 и P_2 разделены расстоянием d, соотношение

$$I(\varphi, d)=2\left(\frac{2J_1(u)}{u}\right)^2\left[1+\left|\frac{2J_1(v)}{v}\right|\cos(\beta_{12}(v)-Cuv)\right], \tag{39}$$

где

и

$$\left.\begin{aligned}\beta_{12}(v)&=0, \text{ когда } \frac{2J_1(v)}{v}>0,\\ \beta_{12}(v)&=\pi, \text{ когда } \frac{2J_2(v)}{v}<0.\end{aligned}\right\} \tag{40}$$

На рис. 10.6 показаны фотографии картин, полученных с помощью такого устройства для различных расстояний d. Приведены также соответствующие теоретические кривые, рассчитанные по формуле (39). Пунктиром показаны огибающие

$$\left.\begin{aligned}I_{\text{макс}}(\varphi, d)&=2\left(\frac{2J_1(u)}{u}\right)^2\left\{1+\left|\frac{2J_1(v)}{v}\right|\right\},\\ I_{\text{мин}}(\varphi, d)&=2\left(\frac{2J_1(u)}{u}\right)^2\left\{1-\left|\frac{2J_1(v)}{v}\right|\right\}.\end{aligned}\right\} \tag{41}$$

Интересно отметить, что при $\beta=\pi$ (случаи (*Г*) и (*Д*)) интенсивность в центре каждой картины имеет в соответствии с нашими общими выводами относительный минимум, а не максимум. Изменение степени когерентности при взаимном удалении обоих отверстий показано на рис. 10.7. Соответствующими буквами там указаны те шесть положений, к которым относятся фотографии на рис. 10.6.

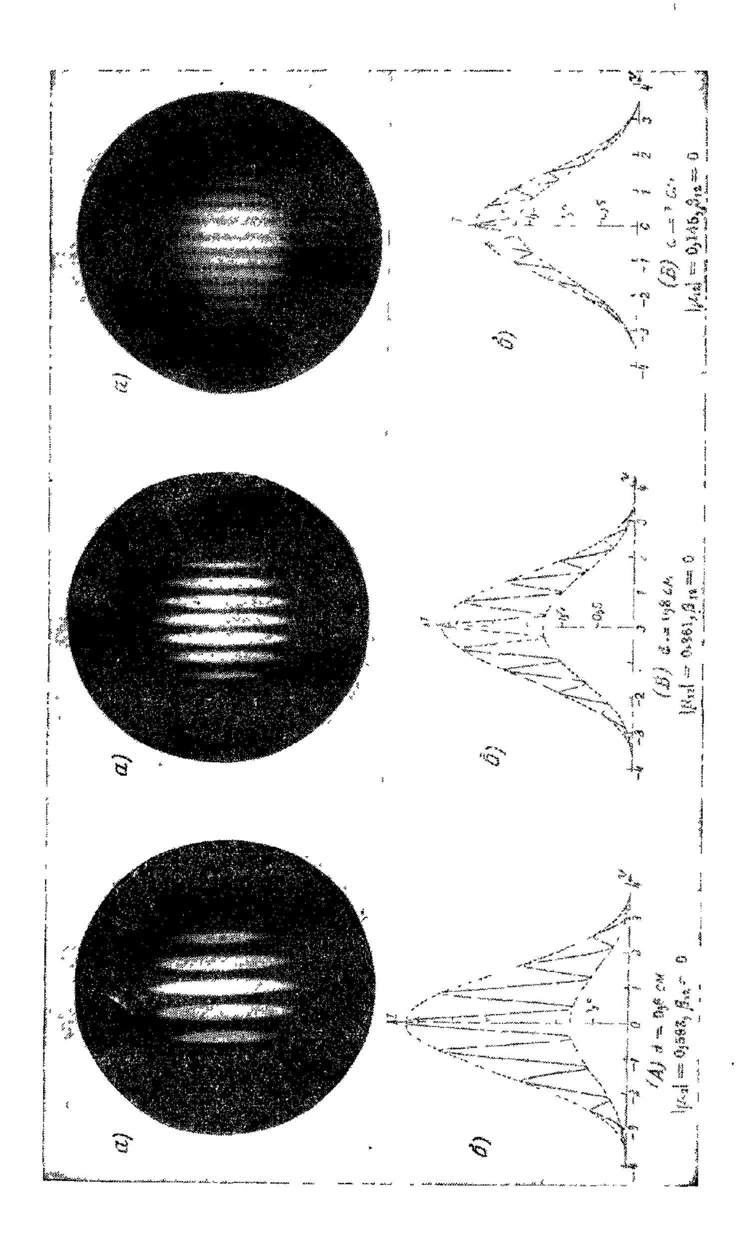
10.4.4. Распространение взаимной интенсивности. Рассмотрим пучок квазимонохроматического света от протяженного первичного источника σ и предположим, что взаимная интенсивность известна для любой пары точек на воображаемой поверхности $\mathcal{A}$, пересекающей пучок. Мы покажем, что в этом случае можно определить взаимную интенсивность для каждой пары точек на любой другой поверхности $\mathcal{B}$, освещаемой светом от $\mathcal{A}$ либо непосредственно, либо через оптическую систему.

Предположим, что среда между $\mathcal{A}$ и $\mathcal{B}$ однородна и ее показатель преломления равен единице. Пусть $U(S, Q_1)$ и $U(S, Q_2)$ — возмущения в точках Q_1 и Q_2 на поверхности $\mathcal{B}$ (рис. 10.8), обусловленные произвольной точкой S ассоциированного монохроматического источника. Тогда, согласно (35), взаимная интенсивность $J(Q_1, Q_2)$ определяется выражением

$$J(Q_1, Q_2)=\int_\sigma U(S, Q_1)U^*(S, Q_2)\,dS. \tag{42}$$

Величины $U(S, Q_1)$ и $U(S, Q_2)$ на основании принципа Гюйгенса — Френеля можно выразить через возмущения во всех точках поверхности $\mathcal{A}$ в виде

$$U(S, Q_1)=\int_{\mathcal{A}} U(S, P_1)\frac{\exp(i\bar{k}s_1)}{s_1}\Lambda_1\,dP_1. \tag{43}$$



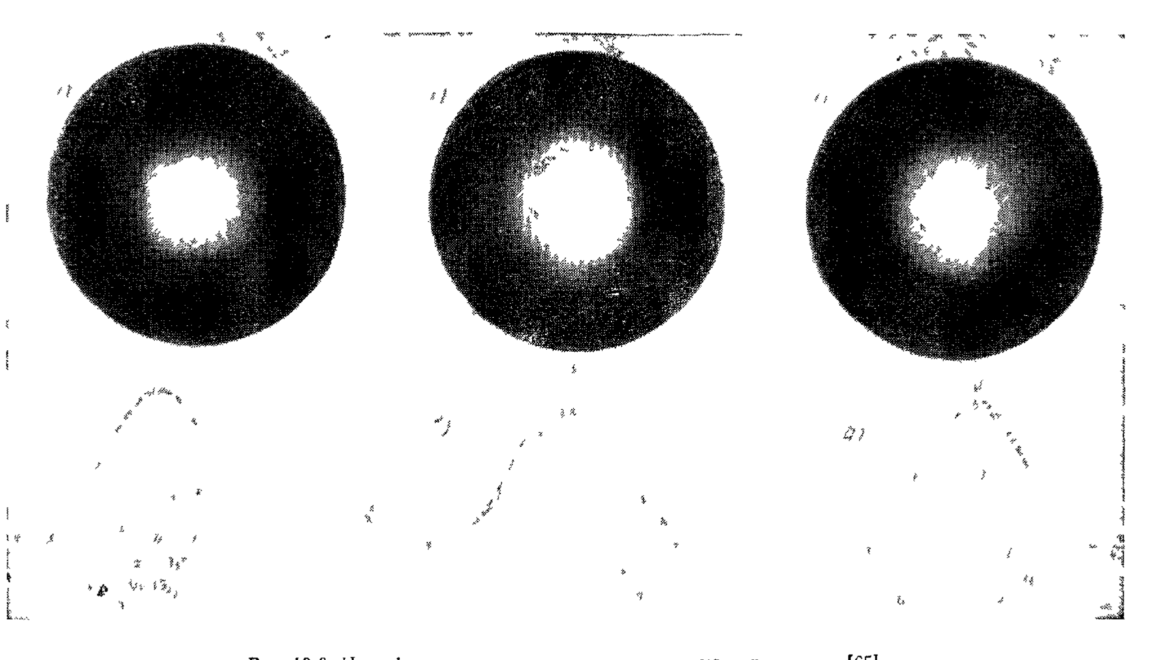


Рис. 10.6. Интерференция двух пучков частично когерентного света [65].

а – наблюдаемые картины; *б* – теоретические кривые интенсивности. Фокусные расстояния линз L_0, L_1 и L_2 дифрактометра $f_0=20$ *см*, $f_1=f_2=R=152$ *см*. Диаметр $L_0=5$ *см*. Расстояние от L_0 до σ_1 составляет 40 *см*, расстояние между L_1 и L_2 – 14 *см*, расстояние от зеркала M до L_2 – 85 *см*. Диаметр $2\rho_1$ отверстия σ_1 равен $0{,}9\cdot10^{-2}$ *см*. Диаметр $2a$ отверстий в точках P_1 и P_2 – 0,14 *см*. Средняя длина волны $\bar{\lambda}=5790$ Å.

Здесь s_1 — расстояние между точкой Q_1 и произвольной точкой P_1 на поверхности $\mathcal{A}$, Λ_1 — коэффициент наклона в точке P_1 (в гл. 8 он обозначался буквой K), $\bar{k}=2\pi\bar{\nu}/c$ — среднее волновое число. Для небольших отклонений от нормали к поверхности $\mathcal{A}$ $\Lambda\approx -i/\bar{\lambda}$. Используя (43) и аналогичное выражение для $U(Q_2)$, получим

$$U(S,\ Q_1)U^*(S,\ Q_2)=\iint\limits_{\mathcal{A}\mathcal{A}} U(S,P_1)\,U^*(S,P_2)\,\frac{\exp\left[i\bar{k}(s_1-s_2)\right]}{s_1\cdot s_2}\,\Lambda_1\Lambda_2^*\,dP_1\,dP_2, \quad (44)$$

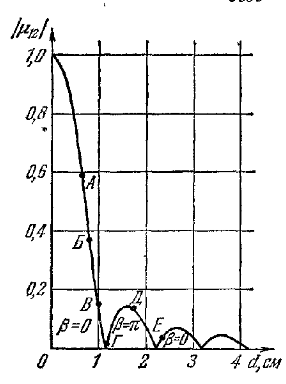


Рис. 10.7. Интерференция двух пучков частично когерентного света.

Степень когерентности как функция расстояния d между двумя освещаемыми отверстиями в дифрактометре $\rho_1=0{,}45\cdot10^{-2}$ см, $R=152$ см, $\bar{\lambda}=5790$ Å, предполагается некогерентное освещение σ_1.

причем точки P_1 и P_2 независимо друг от друга пробегают всю поверхность интегрирования $\mathcal{A}$. Подставим теперь (44) в (42) и изменим порядок интегрирования. Интегрирование по σ дает точно $J(P_1,\ P_2)$, и мы получим

$$J(Q_1,\ Q_2)=\iint\limits_{\mathcal{A}\mathcal{A}} J(P_1,\ P_2)\frac{\exp\left[i\bar{k}(s_1-s_2)\right]}{s_1\cdot s_2}\Lambda_1\Lambda_2^*\,dP_1\,dP_2. \quad (45)$$

Эта формула, предложенная Цернике [66], описывает распространение взаимной интенсивности. При ее выводе мы неявно предполагали, что свет от каждой точки поверхности $\mathcal{A}$ достигает точек Q_1 и Q_2. Наличие любой диафрагмы между обеими поверхностями можно учесть, если ограничиться интегрированием лишь по тем частям поверхности $\mathcal{A}$, которые посылают свет к Q_1 и Q_2. Такой способ приводит к неверному результату, если диафрагма столь мала, что нельзя пренебречь дифракционными эффектами на ее краях. Дифракцию можно учесть, выполняя переход от $\mathcal{A}$ к $\mathcal{B}$ в два этапа, сначала от $\mathcal{A}$ к плоскости диафрагмы, а затем от плоскости диафрагмы к поверхности $\mathcal{B}$.

В специальном случае, когда точки Q_1 и Q_2 совпадают, (45) переходит в следующее выражение для интенсивности:

$$I(Q)=\iint\limits_{\mathcal{A}\mathcal{A}} \sqrt{I(P_1)}\sqrt{I(P_2)}\ \mu(P_1,\ P_2)\,\frac{\exp\left[i\bar{k}(s_1-s_2)\right]}{s_1 s_2}\,\Lambda_1\Lambda_2^*\,dP_1\,dP_2. \quad (46)$$

Здесь мы выразили $J(P_1,\ P_2)$ через интенсивности $I(P_1)$, $I(P_2)$ и комплексную

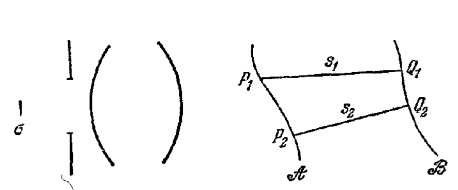


Рис. 10.8. Распространение взаимной интенсивности. К выводу формулы (45).

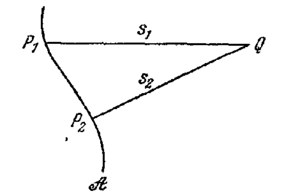


Рис. 10.9. К выводу формулы (46).

степень когерентности $\mu(P_1,\ P_2)$. В этой формуле интенсивность в точке Q представлена в виде суммы вкладов от каждой пары элементов dP_1, dP_2 произвольной поверхности $\mathcal{A}$, пересекающей пучок (рис. 10.9). Вклад от каждой

пары элементов зависит от интенсивности в точках P_1 и P_2 и содержит в качестве весового множителя соответствующее значение комплексной степени когерентности $\mu(P_1, P_2)$. Соотношение (46) можно считать выражением принципа Гюйгенса — Френеля для распространения интенсивности в частично когерентном поле. Сходство только что выведенных формул с формулами более элементарной теории Гюйгенса — Френеля имеет глубокий смысл, который выяснится при строгой формулировке теории частичной когерентности (см. ниже, § 10.7).

Если на пути света от $\mathcal{A}$ до $\mathcal{B}$ находится оптическая система, нужно, очевидно, заменить множитель $(\Lambda \exp(i\bar{k}s))/s$ соответствующей функцией пропускания $K(P, Q)$. Тогда вместо (45) мы получим более общую формулу, а именно

$$J(Q_1, Q_2)=\iint_{\mathcal{A}\mathcal{A}} J(P_1, P_2) K(P_1, Q_1) K^*(P_2, Q_2)\, dP_1\, dP_2. \qquad (47)$$

§ 10.5. Некоторые приложения

10.5.1. Степень когерентности в изображении протяженного некогерентного квазимонохроматического источника. Прежде чем переходить к изучению образования изображения в частично когерентном свете, полезно рассчитать степень когерентности в изображении протяженного некогерентного источника, полученном с помощью центрированной оптической системы. Конечная степень корреляции между колебаниями в плоскости изображения возникает из-за того, что свет от каждой точки источника не собирается в одну точку, а распределяется по конечной площадке, вследствие дифракции (а в общем случае и аберраций). Некоторые из таких «площадок» перекрываются, поэтому в интенсивности в достаточно близких друг к другу точках в плоскости изображения вносится как когерентный, так и некогерентный вклад.

Предположим, что σ — однородный квазимонохроматический некогерентный источник в виде круга радиуса ρ, излучающий свет со средней (в вакууме) длиной волны $\bar{\lambda}_0$ и расположенный в пространстве предмета в однородной среде с показателем преломления n. Пусть далее D — расстояние между плоскостью предмета и плоскостью входного зрачка. Соответствующие величины в пространстве изображения мы будем обозначать теми же символами со штрихом.

Обозначим через d расстояние между двумя точками P_1 и P_2 входного зрачка. Предположим, что $\rho/D \ll 1$, $d/D \ll 1$ и $OP_1 - OP_2 \ll \bar{\lambda}_0$, где O — центральная точка источника *). В этом случае комплексная степень когерентности $\mu(P_1, P_2)$ в соответствии с (10.4.28) равна

$$\mu(P_1, P_2)=\frac{2J_1(v)}{v}, \qquad (1)$$

$$v=\frac{2\pi}{\bar{\lambda}} d \sin\alpha = \frac{2\pi n}{\bar{\lambda}_0} d \sin\alpha, \qquad (2)$$

где $\alpha \approx \sin\alpha \approx \rho/D$ — угол, под которым радиус источника виден из центра входного зрачка (рис. 10.10).

Комплексную степень когерентности для любой пары точек в плоскости выходного зрачка мы могли бы определить, применяя закон распространения (10.4.47). Однако здесь мы рассматриваем специальный случай распространения от одной плоскости к сопряженной ей плоскости, и этот закон принимает более простую форму, которую можно получить следующим образом.

Пусть $U(S, P_1)$ и $U(S, P_2)$ — комплексные возмущения в точках P_1 и P_2, обусловленные возмущением в точке S ассоциированного монохроматического

*) Величина $|\mu(P_1, P_2)|$ не изменится, если последнее условие не выполняется, но фаза $\mu(P_1, P_2)$, согласно (10.4.28), увеличится на величину $\psi - 2\pi\,[OP_1 - OP_2]/\bar{\lambda}_0$.

источника (см. стр. 473); тогда возмущения, обусловленные возмущением в точке S, в точках выходного зрачка, сопряженных P_1 и P_2, определяются формулами

$$U(S,\ P_1') = K_{11}U(S,\ P_1) \qquad U(S,\ P_2') = K_{22}U(S,\ P_2). \tag{3}$$

Здесь $K_{11} = K(P_1,\ P_1')$ — соответствующая функция пропускания для распространения возмущений между сопряженными точками P_1 и P_1' плоскостей зрачков. По формуле Гопкинса (10.4.35б) имеем

$$\mu(P_1',\ P_2') = \frac{1}{\sqrt{I(P_1')}\sqrt{I(P_2')}} \int\limits_\sigma U(S,\ P_1')\,U^*(S,\ P_2')\,dS. \tag{4}$$

Интенсивности $I(P_1')$ и $I(P_1)$ в плоскостях обоих зрачков связаны соотношением

$$I(P_1') = \int\limits_\sigma |U(S,\ P_1')|^2\,dS = |K_{11}|^2 \int\limits_\sigma |U(S,\ P_1)|^2\,dS = |K_{11}|^2\,I(P_1). \tag{5}$$

Аналогичным соотношением связаны $I(P_2')$ и $I(P_2)$. Из (3), (4) и (5) получим

$$\mu(P_1',\ P_2') = \frac{K_{11}K_{22}^*}{|K_{11}|\,|K_{22}|} \cdot \frac{1}{\sqrt{I(P_1)}\sqrt{I(P_2)}} \int\limits_\sigma U(S,\ P_1)\,U^*(S,\ P_2)\,dS =$$
$$= \exp[i(\Phi_{11} - \Phi_{22})]\,\mu(P_1,\ P_2), \tag{6}$$

где Φ_{11} и Φ_{22} — фазы K_{11} и K_{22} соответственно. Это соотношение означает, что

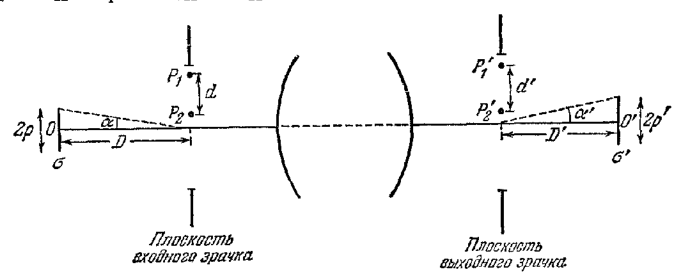


Рис. 10.10. К вычислению степени когерентности в изображении некогерентного источника света.

степень когерентности $|\mu|$ *для любых двух точек выходного зрачка равна степени когерентности для сопряженных точек входного зрачка, а фазы соответствующих значений комплексных степеней когерентности для соответствующих пар точек отличаются на величину* $\Phi_{11} - \Phi_{22}$, *т. е. на геометрическую разность фаз* $2\pi\{[P_1P_1'] - [P_2P_2']\}/\lambda_0$. Пусть

$$v' = \frac{2\pi n'}{\lambda_0}\,d' \sin\alpha'. \tag{7}$$

Так как точка P_1' сопряжена с P_1, а точка P_2' — с P_2, то в рамках параксиальной оптики из теоремы Смита — Гельмгольца (см. (4.4.49)) следует, что *) $v' = v$. Следовательно, комплексную степень когерентности для пары точек в выходном зрачке можно, согласно (1) и (6), записать в виде

$$\mu(P_1',\ P_2') = \left(\frac{2J_1(v')}{v'}\right) \exp[i(\Phi_{11} - \Phi_{22})]. \tag{8}$$

Если, как и в (10.4.31), считать значения $|\mu| \geqslant 0{,}88$ достаточно хорошими приближениями к полной когерентности и вспомнить, что $|2J(v)/v| \geqslant 0{,}88$, когда

*) Это означает, что v и v' соответствуют особому выбору переменных Зайделя (см. § 5.2).

$v \leqslant 1$, то окажется, что *некогерентный квазимонохроматический однородный круглый источник будет когерентно освещать в выходном зрачке площадки диаметром*

$$d'_{\text{ког}} \approx \frac{0,16\bar{\lambda}_0}{n' \sin \alpha'}, \tag{9}$$

где $2\alpha' \approx 2\rho'/D'$ — угол, под которым диаметр изображения источника виден из центра выходного зрачка и $\bar{\lambda}_0/n' = \bar{\lambda}'$ — средняя длина волны света в пространстве изображения.

Запишем (9) в несколько иной форме. Пусть r'_A обозначает радиус первого темного кольца в картине Эйри, связанной с системой, т. е.

$$r'_A = \frac{0,61\bar{\lambda}_0}{n' \sin \theta'}, \tag{10}$$

где $n' \sin \theta' \approx n'a'/D'$ — числовая апертура со стороны изображения. Тогда, согласно (9) и (10), $d'_{\text{ког}}/r'_A \approx 0,16 \sin \theta'/0,61 \sin \alpha'$, и, следовательно,

$$d'_{\text{ког}} \approx 0,26\, a' \left(\frac{r'_A}{\rho'}\right). \tag{11}$$

В этой формуле размеры когерентно освещаемых областей выходного зрачка выражены через «физические параметры», а именно: радиус r'_A первого темного кольца соответствующей картины Эйри, радиус ρ' геометрического изображения источника и радиус a' выходного зрачка.

Выходной зрачок и, следовательно, плоскость изображения освещаются почти *когерентно*, если $d'_{\text{ког}} \gg a'$, т. е. если

$$\rho' \leqslant 0,26 r'_A. \tag{12}$$

Когда $d'_{\text{ког}} \ll a'$, т. е. когда

$$\rho' \gg 0,26 r'_A, \tag{13}$$

когерентно освещаемые площадки выходного зрачка будут малы по сравнению с самим выходным зрачком, так что в этом случае выходной зрачок фактически освещается *некогерентно*. При этом комплексная степень когерентности для пар точек Q'_1, Q'_2 в плоскости изображения будет совпадать со степенью когерентности, обусловленной некогерентным источником, расположенным там же, где и выходной зрачок, и обладающим теми же размерами и формой, причем распределение интенсивности вдоль источника и выходного зрачка одинаково. Следовательно, согласно теореме Ван-Циттерта — Цернике (10.4.21), имеем

$$\mu(Q'_1, Q'_2) = \frac{1}{\sqrt{I(Q'_1)}\sqrt{I(Q'_2)}} \int_{\mathcal{A}'} I(P') \frac{\exp[i\bar{k}(s_1 - s_2)]}{s_1 s_2} dP', \tag{14}$$

$$I(Q'_1) = \int_{\mathcal{A}'} \frac{I(P')}{s_1^2} dP', \qquad I(Q'_2) = \int_{\mathcal{A}'} \frac{I(P')}{s_2^2} dP'. \tag{15}$$

Интегрирование проводится по выходному зрачку $\mathcal{A}'$, а через s_1 и s_2 обозначают расстояния от произвольной точки P' на $\mathcal{A}'$ до точек Q'_1 и Q'_2, соответственно (рис. 10.11, *а*). Интенсивность $I(P')$ можно рассчитать, зная интенсивность $I(P)$ в сопряженной точке входного зрачка, посредством соотношения (5). Поскольку в это соотношение не входит фаза функции пропускания, величина $\mu(Q'_1, Q'_2)$ *не зависит от аберраций системы*. Как правило, интенсивность $I(P')$ практически постоянна; кроме того, если точки Q'_1 и Q'_2 достаточно близки друг к другу, выражение (14) принимает вид

$$\mu(Q'_1, Q'_2) = \frac{2J_1(u')}{u'}, \qquad u' = \frac{2\pi n'}{\bar{\lambda}_0} \frac{a'}{D'} h', \tag{16}$$

где h' — расстояние между Q'_1 и Q'_2.

В общем случае, когда не выполняется ни условие (12), ни условие (13), выходной зрачок освещается частично когерентным светом, который характеризуется комплексной степенью когерентности (8). Тогда значение комплексной степени когерентности для пары точек в плоскости изображения нужно рассчитывать с помощью закона распространения (10.4.45), что приводит к выражению

$$\mu(Q_1', Q_2')=\frac{1}{\sqrt{I(Q_1')}\sqrt{I(Q_2')}}\iint\limits_{\mathcal{A}'\mathcal{A}'}\sqrt{I(P_1')}\sqrt{I(P_2')}\left(\frac{2J_1(v')}{v'}\right)\times$$
$$\times\frac{\exp\{i[\Phi_{11}-\Phi_{22}+\bar{k}(s_1-s_2)]\}}{s_1s_2}\Lambda_1\Lambda_2^*\,dP_1'dP_2'. \quad (17)$$

Пользуясь им, можно также рассчитать интенсивности $I(Q_1')$ и $I(Q_2')$, если

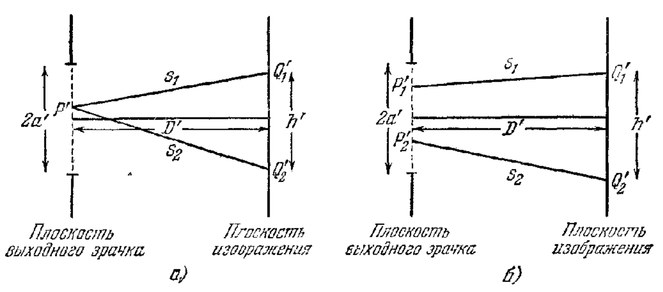


Рис 10.11. К вычислению комплексной степени когерентности в плоскости изображения. а — некогерентное освещение выходного зрачка; б — частично когерентное освещение выходного зрачка.

учесть, что $\mu(Q_1', Q_1')=\mu(Q_2', Q_2')=1$. Отметим, что в этом случае комплексная степень когерентности зависит от аберраций системы, поскольку подынтегральное выражение содержит фазы Φ_{11} и Φ_{22} функции пропускания.

10.5.2. Влияние конденсора на разрешающую силу микроскопа. Для того чтобы исследовать под микроскопом небольшой несветящийся объект, его необходимо осветить. Если объект почти прозрачен, как это обычно бывает, он освещается снизу, т. е. освещается *проходящим светом*, и свет, прошедший через

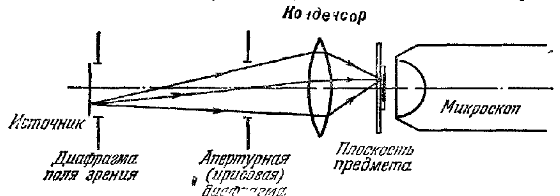


Рис. 10.12. Критическое освещение.

объект, фокусируется в плоскости изображения объектива микроскопа. Для получения достаточной освещенности обычно используют вспомогательную систему линз — конденсор. Существуют различные методы освещения. Здесь мы кратко опишем два широко распространенных метода, так называемое *критическое освещение* и *освещение по Келеру*, и рассмотрим разрешающую силу, которую можно достичь с их помощью.

а. *Критическое освещение.* В этом методе равномерно яркий источник располагается непосредственно за диафрагмой поля зрения и с помощью конденсора изображается на плоскость предмета объектива микроскопа (рис. 10.12).

Размер диафрагмы поля зрения подбирается так, чтобы ее изображение конденсором точно покрывало поле зрения.

Размер освещенной области в плоскости изображения конденсора (плоскость предмета объектива) значительно больше, чем эффективный размер диска Эйри, создаваемого одной точкой источника (в обозначениях, принятых в п. 10.5.1, $\rho' \gg r'_A$). Согласно п. 10.5.1 при таких условиях *комплексная степень когерентности для любой пары точек в плоскости предмета объектива совпадает со степенью когерентности, обусловленной некогерентным источником, заполняющим конденсор.* Кроме того, степень когерентности не зависит от аберраций конденсора. Очевидно, что разрешающая сила микроскопа зависит только от степени когерентности света, падающего на предмет и от свойств объектива. Следовательно, *аберрации конденсора совершенно не влияют на разрешающую силу микроскопа.* Этот важный результат, впервые полученный другим способом Цернике [67], показывает ошибочность широко распространенного мнения, согласно которому хорошо скорректированный конденсор обладает преимуществами при получении высокой разрешающей силы.

Для оценки влияния размеров конденсора на разрешение рассмотрим два небольших отверстия $P_1(X_1, Y_1)$ и $P_2(X_2, Y_2)$ в плоскости предмета. Комплексная степень когерентности света, достигающего этих отверстий, при тех же предположениях, что и прежде, определяется формулой типа (16), а именно

$$\mu(P_1, P_2) = \frac{2J_1(u_{12})}{u_{12}}, \qquad u_{12} = \frac{2\pi}{\bar{\lambda}_0}\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}\, n'_c \sin\theta'_c, \tag{18}$$

где $n'_c \sin\theta'_c$ — числовая апертура конденсора со стороны объектива микроскопа. Пусть $P(X, Y)$ — любая другая точка в плоскости предмета, а P' — ее изображение объективом. Предположим, что объектив практически свободен от аберраций. Тогда распределение интенсивности в плоскости изображения объектива для света, приходящего только от P_1, представляет собой картину Эйри, центром которой служит изображение P'_1 точки P_1. Следовательно, если $n_0 \sin\theta_0$ — числовая апертура объектива, то интенсивность $I^{(1)}(P')$, обусловленная светом, который приходит в точку P' только из P_1, с точностью до постоянного множителя равна

$$I^{(1)}(P') = \left(\frac{2J_1(v_1)}{v_1}\right)^2, \qquad v_1 = \frac{2\pi}{\bar{\lambda}_0}\sqrt{(X-X_1)^2+(Y-Y_1)^2}\, n_0 \sin\theta_0. \tag{19a}$$

Интенсивность $I^{(2)}(P')$, обусловленная светом, который попадает в эту точку от отверстия P_2, определяется аналогичным выражением

$$I^{(2)}(P') = \left(\frac{2J_1(v_2)}{v_2}\right)^2, \qquad v_2 = \frac{2\pi}{\bar{\lambda}_0}\sqrt{(X-X_2)^2+(Y-Y_2)^2}\, n_0 \sin\theta_0. \tag{19б}$$

Таким образом, если два отверстия освещаются через конденсор, то интенсивность $I(P')$ в плоскости изображения объектива микроскопа обусловливается суперпозицией двух частично когерентных пучков. Интенсивность каждого из пучков определяется выражениями (19), а комплексная степень когерентности — соотношением (18). Подставляя указанные соотношения в формулу (10.4.11), сразу же получим выражение для $I(P')$. Если предположить также, что точка P' достаточно близка к геометрическим изображениям точек P_1 и P_2 (точнее, что $\delta = [P_1P'] - [P_2P'] \ll \bar{\lambda}$), то формула для $I(P')$ примет вид

$$I(P') = \left(\frac{2J_1(v_1)}{v_1}\right)^2 + \left(\frac{2J_1(v_2)}{v_2}\right)^2 + 2\left(\frac{2J_1(mv_{12})}{mv_{12}}\right)\left(\frac{2J_1(v_1)}{v_1}\right)\left(\frac{2J_1(v_2)}{v_2}\right), \tag{20}$$

где

$$m = \frac{n'_c \sin\theta'_c}{n_0 \sin\theta_0}, \qquad v_{12} = \frac{u_{12}}{m} = \frac{2\pi}{\bar{\lambda}_0}\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}\, n_0 \sin\theta_0. \tag{21}$$

Из соотношения (20) можно вывести ряд интересных заключений. Если

mv_{12} является отличным от нуля корнем уравнения $J_1(mv_{12})=0$, то в соотношении (20) отсутствует член, содержащий произведения, и оно принимает вид

$$I(P')=\left(\frac{2J_1(v_1)}{v_1}\right)^2+\left(\frac{2J_1(v_2)}{v_2}\right)^2. \qquad (22)$$

Распределение интенсивности в плоскости изображения оказывается тогда таким же, как и при *некогерентном* освещении отверстий P_1 и P_2. Это осуществляется, например, когда $m=1$, а v_{12}— отличный от нуля корень уравнения $J_1(v_{12})=0$, т. е. когда числовые апертуры равны, а расстояние между геометрическими изображениями отверстий равно радиусу одного из темных колец картины Эйри, создаваемой объективом.

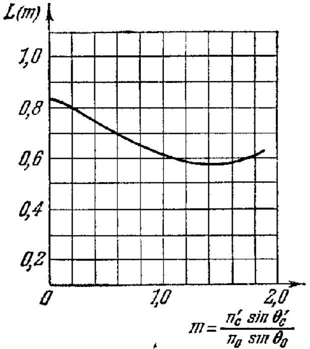


Рис. 10.13. Влияние апертуры конденсора на разрешение изображений двух небольших отверстий равной яркости [68].

Если числовая апертура конденсора очень мала ($m\to 0$), то $2J(mv_{12})/mv_{12}\approx 1$ и (20) сводится к

$$I(P')=\left\{\frac{2J_1(v_1)}{v_1}+\frac{2J_1(v_2)}{v_2}\right\}^2. \qquad (23)$$

При этом для любого расстояния между отверстиями распределение интенсивности остается таким же, как и в случае полностью *когерентного* освещения.

Формула (20) позволяет изучить зависимость распределения интенсивности в плоскости изображения объектива микроскопа от отношения числовых апертур m. В частности, определим интенсивность в точке, находящейся посередине между P_1' и P_2'. Будем считать, что изображения отверсти начинают разрешаться, когда интенсивность в этой точке на 26,5% меньше, чем интенсивность в каждой из наших двух точек. Величина 26,5% соответствует критерию Рэлея для круглого отверстия при некогерентном освещении (см. п. 8.6.2). Выразим предельное разделение $(P_1P_2)_{\text{пред}}$, соответствующее этому критерию, в одинаковом виде как для некогерентного (см. (8.6.32)), так и для когерентного (см. (8.6.55)) освещения

$$(P_1P_2)_{\text{пред}}=L(m)\frac{\bar{\lambda}_0}{n_0\sin\theta_0}. \qquad (24)$$

На рис. 10.13 изображена кривая $L(m)$, рассчитанная из (20) на основе указанного критерия. Как мы видим, наилучшее разрешение получается при $m\approx 1,5$, т. е. при числовой апертуре конденсора, примерно в полтора раза превышающей числовую апертуру объектива. Величина L в этом случае несколько меньше значения 0,61, получающегося при некогерентном освещении.

б. Освещение по Кёлеру. В предложенном Кёлером [69] методе освещения, схема которого приведена на рис. 10.14, собирательная линза располагается рядом с диафрагмой поля зрения и образует изображение источника σ в фокальной плоскости конденсора (где расположена диафрагма конденсора). Тогда лучи, исходящие из каждой точки источника, образуют после прохождения конденсора параллельные пучки. Преимущество такого устройства заключается в том, что неравномерность в распределении яркости по источнику не вызывает неравномерности в освещенности поля зрения.

Чтобы оценить предельное разрешение, достигаемое при освещении методом Кёлера, необходимо прежде всего определить комплексную степень когерентности μ для пар точек в плоскости предмета объектива микроскопа. Пусть

$$U(S, P_1)=A_1\exp(i\varphi_1), \qquad U(S, P_2)=A_2\exp(i\varphi_2) \qquad (25)$$

— комплексные возмущения в точках $P_1(X_1, Y_1)$ и $P_2(X_2, Y_2)$ плоскости предмета объектива микроскопа, обусловленные возмущением в точке S монохроматического источника, ассоциированного с σ. Очевидно,

$$\varphi_1-\varphi_2=\frac{2\pi}{\lambda_0}\left[p(X_1-X_2)+q(Y_1-Y_2)\right], \tag{26}$$

где p и q — первые две лучевые компоненты двух параллельных лучей, испускаемых точкой S источника и проходящих через точки P_1 и P_2. При наличии аберраций в конденсорной системе два луча не будут строго параллельны, однако такой непараллельностью вполне можно пренебречь, поскольку мы

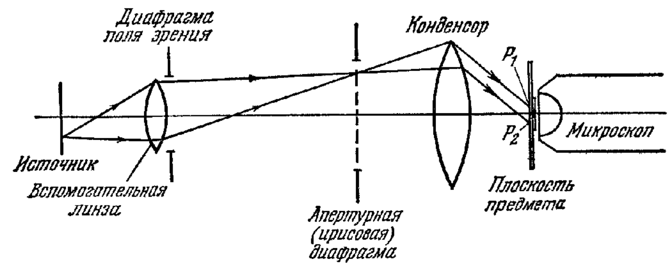


Рис. 10.14. Освещение по Кёллеру.

рассматриваем лишь точки, близкие друг другу. Подставляя (25) и (26) в формулу Гопкинса (10.4.35б), получим

$$\left.\begin{aligned}&\mu(P_1, P_2)=\\&=\frac{1}{\sqrt{I(P_1)}\sqrt{I(P_2)}}\int_\sigma A_1A_2\exp\left\{i\bar{k}_0\left[p(X_1-X_2)+q(Y_1-Y_2)\right]\right\}dS,\\&\qquad I_1(P)=\int_\sigma A_1^2\,dS,\qquad I_2=\int_\sigma A_2^2\,dS.\end{aligned}\right\} \tag{27}$$

Так как каждой точке $S(\xi, \eta)$ источника соответствует пара лучевых компонент (p, q), мы можем от интегрирования по σ перейти к интегрированию по телесному углу

$$p^2+q^2\leqslant n_c'^2\sin^2\theta_c', \tag{28}$$

образуемому лучами, падающими на предмет. В приближении параксиальной оптики соотношения $\xi=\xi(p, q)$ и $\eta=\eta(p, q)$ линейны. В самом деле, как легко показать из формул (4.3.10), $\xi=fp$, $\eta=fq$, где f — фокусное расстояние конденсорной системы. Следовательно, якобиан $\partial(\xi, \eta)/\partial(p, q)$ равен постоянной величине. Вне рамок геометрической оптики якобиан в общем случае изменяется по области интегрирования, однако этим изменением можно пренебречь, поскольку оно достаточно мало по сравнению с изменением экспоненциального члена. Если пренебречь также медленным изменением A_1 и A_2, то (27) примет вид

$$\mu(P_1, P_2)=\frac{\iint\limits_\Omega \exp\left\{i\bar{k}_0\left[p(X_1-X_2)+q(Y_1-Y_2)\right]\right\}dp\,dq}{\iint\limits_\Omega dp\,dq}, \tag{29}$$

где Ω означает область (28). Вычисляя (29), получим

$$\mu(P_1, P_2)=\frac{2J_1(u_{12})}{u_{12}},\qquad u_{12}=\frac{2\pi}{\lambda_0}\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}\,n_c'\sin\theta_c'. \tag{30}$$

Эта формула идентична формуле (18) в случае критического освещения. Следовательно, *комплексная степень когерентности света, падающего на плоскость предмета микроскопа, одинакова как при критическом освещении, так и при освещении по Кёлеру.* Этот результат показывает, что часто употребляемые названия «некогерентное» для критического освещения и «когерентное» для освещения по Кёлеру нужно считать неудачными. Как мы видим, формула (20) справедлива для обоих типов освещения, а рис. 10.13 одинаково применим к обоим случаям.

10.5.3. Получение изображения при частично когерентном квазимонохроматическом освещении*). *а. Распространение взаимной интенсивности через оптическую систему.* В § 9.5 было описано несколько общих методов изучения отображения протяженных объектов. Рассматривались случаи полностью когерентного (п. 9.5.1) и полностью некогерентного (п. 9.5.2) освещения. В первом случае рассматривалось распространение через систему комплексной амплитуды, во втором — интенсивности. Сейчас мы исследуем более общий случай частично когерентного квазимонохроматического освещения. Изучаемой величиной здесь является взаимная интенсивность.

Как и в (9.5.1), используем нормализованные координаты Зайделя, так что точка предмета и ее параксиальное изображение имеют одинаковые численные значения координат. Пусть $J_0(x_0, y_0; x_0', y_0')$ — взаимная интенсивность для точек (x_0, y_0), (x_0', y_0') в плоскости предмета. Если $K(x_0, y_0; x_1, y_1)$ — функция пропускания системы (см. п. 9.5.1), то взаимная интенсивность в плоскости изображения, согласно закону распространения (10.4.47), определяется выражением

$$J_1(x_1, y_1; x_1', y_1') = \iiiint_{-\infty}^{+\infty} J_0(x_0, y_0; x_0', y_0') K(x_0, y_0; x_1, y_1) K^*(x_0', y_0'; x_1', y_1') dx_0\, dy_0\, dx_0'\, dy_0'. \tag{31а}$$

Интегрирование лишь формально производится по бесконечной области, так как для всех точек в плоскости предмета, от которых свет не попадает в плоскость изображений, величина J_0 равна нулю.

Как и в § 9.5, мы предположим, что предмет так мал, что служит изопланатической областью системы, т. е. что для всех его точек величину $K(x_0, y_0; x_1, y_1)$ с хорошей точностью можно заменить функцией, зависящей лишь от разностей x_1-x_0 и y_1-y_0 (скажем, $K(x_1-x_0, y_1-y_0)$). Уравнение (31а) тогда примет вид

$$J_1(x_1, y_1; x_1', y_1') = \iiiint_{-\infty}^{+\infty} J_0(x_0, y_0; x_0', y_0') K(x_1-x_0, y_1-y_0) \times K^*(x_1'-x_0', y_1'-y_0')\, dx_0\, dy_0\, dx_0'\, dy_0'. \tag{31б}$$

Представим J_0, J_1 и произведение KK^* в форме четырехмерных интегралов Фурье, а именно

$$J_0(x_0, y_0; x_0', y_0') = \iiiint_{-\infty}^{+\infty} \mathcal{J}_0(f, g; f', g') \exp[-2\pi i(fx_0 + gy_0 + f'x_0' + g'y_0')]\, df\, dg\, df'\, dg', \tag{32а}$$

$$J_1(x_1, y_1; x_1', y_1') = \iiiint_{-\infty}^{+\infty} \mathcal{J}_1(f, g; f', g') \exp[-2\pi i(fx_1 + gy_1 + f'x_1' + g'y_1')]\, df\, dg\, df'\, dg', \tag{32б}$$

*) Содержание настоящего раздела частично основано на исследованиях Гопкинса [12] и Дюмонте [70].

$$K(x, y) K^*(x', y') =$$
$$= \iiiint\limits_{-\infty}^{+\infty} \mathscr{M}(f, g; f', g') \exp[-2\pi i (fx + gy + f'x' + g'y')]\, df\, dg\, df'\, dg'. \qquad (32в)$$

Тогда на основании обратного преобразования Фурье находим

$$\mathscr{J}_0(f, g; f', g') =$$
$$= \iiiint\limits_{-\infty}^{+\infty} J_0(x_0, y_0; x_0', y_0') \exp[2\pi i (fx_0 + gy_0 + f'x_0' + g'y_0')] dx_0\, dy_0\, dx_0'\, dy_0'. \qquad (33)$$

Совершенно аналогичные соотношения можно написать для $\mathscr{J}_1$ и $\mathscr{M}$.

Применяя теорему свертывания к (31б), получим соотношение

$$\mathscr{J}_1(f, g; f', g') = \mathscr{J}_0(f, g; f', g') \mathscr{M}(f, g; f', g'). \qquad (34)$$

Отсюда следует, что если взаимную интенсивность в плоскостях предмета и изображения представить суперпозицией четырехмерных пространственных гармоник всевозможных пространственных частот (f, g, f', g'), то каждая такая компонента взаимной интенсивности в изображении будет зависеть лишь от ее соответствующей компоненты в предмете, а их отношение окажется равным $\mathscr{M}$. Таким образом, в пределах применимости настоящего приближения *влияние оптической системы на взаимную интенсивность эквивалентно действию четырехмерного линейного фильтра*. Функция $\mathscr{M}$ называется *функцией частотного отклика для частично когерентного квазимонохроматического освещения*.

Функция частотного отклика связана с функцией зрачка системы простым соотношением. Если, как и в (9.5.10в), мы представим K в виде двумерного интеграла Фурье

$$K(x, y) = \iint\limits_{-\infty}^{+\infty} \mathscr{K}(f, g) \exp[-2\pi i (fx + gy)]\, df\, dg, \qquad (35)$$

и подставим его в соотношение, обратное преобразованию (32в), то найдем, что

$$\mathscr{M}(f, g; f', g') = \mathscr{K}(f, g) \mathscr{K}^*(-f', -g'). \qquad (36)$$

Согласно (9.5.13) величина $\mathscr{K}(f, g)$ равна значению функции зрачка $G(\xi, \eta)$ системы в точке

$$\xi = \lambda R f, \qquad \eta = \lambda R g \qquad (37)$$

на опорной сфере Гаусса (радиуса R). Следовательно, *для частично когерентного квазимонохроматического освещения функция частотного отклика связана с функцией зрачка системы формулой*

$$\mathscr{M}\left(\frac{\xi}{\lambda R}, \frac{\eta}{\lambda R}, \frac{\xi'}{\lambda R}, \frac{\eta'}{\lambda R}\right) = G(\xi, \eta) G^*(-\xi', -\eta'). \qquad (38)$$

Так как для точек, находящихся вне выходного зрачка, функция зрачка равна нулю, то система не пропустит спектральные компоненты, соответствующие частотам, превышающим определенные значения. Если выходной зрачок имеет форму круга радиуса a, то величина $G(\xi, \eta) G^*(-\xi', -\eta')$ равна нулю при $\xi^2 + \eta^2 > a^2$ или $\xi'^2 + \eta'^2 > a^2$. Следовательно, не будут пропускаться спектральные компоненты взаимной интенсивности, соответствующие частотам (f, g, f', g'), для которых *)

$$f^2 + g^2 > \left(\frac{a}{\bar{\lambda} R}\right)^2 \quad \text{или} \quad f'^2 + g'^2 > \left(\frac{a}{\bar{\lambda} R}\right)^2. \qquad (39)$$

Здесь $\bar{\lambda}$ — средняя длина волны в пространстве изображения.

---

*) Если угловая апертура системы мала и выполняется условие синусов, то, как и в § 9.5, $a/\bar{\lambda} R \approx (n_0 \sin \theta_0)/(M\bar{\lambda}_0)$, где $n_0 \sin \theta_0$ — числовая апертура системы, M — гауссово увеличение и $\bar{\lambda}_0$ — средняя длина волны в вакууме.

В табл. 10.1 приведены основные формулы, относящиеся к отображению при частично когерентном освещении, а также соответствующие формулы из § 9.5 для когерентного и некогерентного освещений. Выражения, относящиеся к некогерентному освещению, можно получить из общих формул (34), (36) и (38), предполагая, что J_0 имеет вид $J_0(x_0, y_0; x_0', y_0') = I(x_0, y_0)\,\delta(x_0' - x_0)\,\delta(y_0' - y_0)$, где δ — дельта-функция Дирака (см. приложение 4). Мы не будем приводить здесь этот вывод, поскольку он, хотя и достаточно прост, но довольно громоздок.

Таблица 10.1

Отклик оптической системы на пространственные частоты.
(Предполагается, что предмет служит изопланатической областью системы.)

Освещение	Основная величина	Переход от предмета к изображению	Функция частотного отклика
Когерентное	Комплексное возмущение $U(x, y)$	$\mathcal{U}_1(f, g) = \mathcal{U}_0(f, g)\,\mathcal{K}(f, g)$	$\mathcal{K}(f, g)$
Некогерентное	Интенсивность $I(x, y)$	$\mathcal{J}_1(f, g) = \mathcal{J}_0(f, g)\,\mathcal{L}(f, g)$	$\mathcal{L}(f, g) = \int\limits_{-\infty}^{+\infty}\int \mathcal{K}(f'+f, g'+g) \times \mathcal{K}^*(f', g')\,df'dg'$
Частично когерентное	Взаимная интенсивность $J(x, y; x', y')$	$\mathcal{Y}_1(f, g; f', g') = \mathcal{Y}_0(f, g; f', g') \times \mathcal{M}(f, g; f', g')$	$\mathcal{M}(f, g; f', g') = \mathcal{K}(f, g) \times \mathcal{K}^*(-f', -g')$

$\mathcal{K}(f, g)$ — обратное преобразование Фурье от функции пропускания $K(x, y)$ системы, связанное с функцией зрачка $G(\xi, \eta)$ формулой $\mathcal{K}(\xi/\lambda R, \eta/\lambda R) = G(\xi, \eta)$, где R — радиус опорной сферы Гаусса и λ — средняя длина волны в пространстве изображения.

Формулы для специального случая идеально монохроматического (и, следовательно, полностью когерентного) освещения можно вывести несколько проще, так как в этом случае имеем для взаимной интенсивности $J_0(x_0, y_0; x_0', y_0') = U(x_0, y_0)\,U_0^*(x_0', y_0')$. Величина $\mathcal{Y}_0$, являющаяся Фурье-образом J_0, также приобретает вид произведения двух множителей. Применяя далее выражения (34) и (36), легко показать, что каждая спектральная компонента комплексного возмущения U_0 распространяется через систему в соответствии с формулами, приведенными в первом ряду табл. 10.1.

б. *Изображения объектов, освещаемых проходящим светом.* Предположим, что часть плоскости предмета занята прозрачным или полупрозрачным объектом, который освещается частично когерентным квазимонохроматическим светом. Будем считать, что этот свет испускается первичным источником и достигает плоскости предмета, проходя через некоторую осветительную систему (конденсор).

Как и в п. 8.6.1, введем для объекта соответствующую функцию пропускания $F_0(x_0, y_0)$. Если через $U_0^-(S; x_0, y_0)$ обозначить возмущение в точке (x_0, y_0) плоскости предмета, обусловленное возмущением в точке S ассоциированного монохроматического источника (см. стр. 471), то после прохождения через предмет возмущение, создаваемое этой точкой источника, определяется

соотношением

$$U_0(S;\ x_0,\ y_0)=U_0^-(S;\ x_0,\ y_0)\,F(x_0,\ y_0). \tag{40}$$

Взаимная интенсивность света, падающего на предмет, согласно (10.4.35а), запишется в виде

$$J_0^-(x_0,y_0;\ x_0',\ y_0')=\int_\sigma U_0^-(S;\ x_0,\ y_0)\,U_0^{-*}(S;\ x_0',\ y_0')\,dS, \tag{41а}$$

а взаимная интенсивность света, выходящего из предмета, в виде

$$J_0(x_0,\ y_0;\ x_0',\ y_0')=\int_\sigma U_0(S;\ x_0,\ y_0)\,U_0^*(S;\ x_0',\ y_0')\,dS; \tag{41б}$$

следовательно, в соответствии с (40) имеем

$$J_0(x_0,\ y_0;\ x_0',\ y_0')=F(x_0,\ y_0)\,F^*(x_0',\ y_0')\,J_0^-(x_0,\ y_0;\ x_0',\ y_0'). \tag{42}$$

Мы ограничимся важным случаем зависимости взаимной интенсивности J_0^- падающего света лишь от разностей x_0-x_0', y_0-y_0' координат x_0, y_0, x_0', y_0', т. е. случаем, когда J_0^- имеет вид

$$J_0^-(x_0,\ y_0;\ x_0',\ y_0')=J_0^-(x_0-x_0',\ y_0-y_0'). \tag{43}$$

В п. 10.5.2 было показано, что это справедливо как для критического освещения, так и для освещения по Кёлеру. По-прежнему будем считать предмет настолько малым, что он служит изопланатической областью системы. Тогда из (31б) следует, что интенсивность $I_1(x_1,\ y_1)=J_1(x_1,\ y_1;\ x_1,\ y_1)$ в плоскости изображения определяется выражением

$$I_1(x_1,\ y_1)=\iiiint\limits_{-\infty}^{+\infty} J_0^-(x_0-x_0',\ y_0-y_0')\,F(x_0,\ y_0)\,F^*(x_0',\ y_0')\times$$
$$\times K(x_1-x_0,\ y_1-y_0)\,K^*(x_1-x_0',\ y_1-y_0')\,dx_0\,dy_0\,dx_0'\,dy_0'. \tag{44}$$

Представим F и J_0^- в форме двумерных интегралов Фурье

$$F(x,\ y)=\iint\limits_{-\infty}^{+\infty}\mathcal{F}(f,\ g)\exp\left[-2\pi i\,(fx+gy)\right]df\,dg, \tag{45а}$$

$$J_0^-(x,\ y)=\iint\limits_{-\infty}^{+\infty}\mathcal{Y}_0^-(f,\ g)\exp\left[-2\pi i\,(fx+gy)\right]df\,dg. \tag{45б}$$

Если мы подставим F и F^* из (45а) в (44), используем тождество $f'x_0-f''x_0'=(f'-f'')\,x_1-f'(x_1-x_0)+f''(x_1-x_0')$, аналогичное тождество для g и y и введем новые переменные интегрирования $u'=x_1-x_0$, $u''=x_1-x_0'$, то получим следующее выражение для I_1:

$$I_1(x_1,y_1)=\iiiint\limits_{-\infty}^{+\infty}\mathcal{T}(f',\ g';\ f'',\ g'')\,\mathcal{F}(f',\ g')\,\mathcal{F}^*(f'',\ g'')\times$$
$$\times\exp\left\{-2\pi i\left[(f'-f'')\,x_1+(g'-g'')\,y_1\right]\right\}df'\,dg'\,df''\,dg'', \tag{46}$$

где

$$\mathcal{T}(f',\ g';\ f'',\ g'')=\iiiint\limits_{-\infty}^{+\infty} J_0^-(u''-u',\ v''-v')\,K(u',\ v')\,K^*(u'',\ v'')\times$$
$$\times\exp\left\{2\pi i\left[(f'u'+g'v')-(f''u''+g''v'')\right]\right\}du'\,dv'\,du''\,dv''=$$
$$=\iiiiiint\limits_{-\infty}^{+\infty}\mathcal{Y}_0^-(f,\ g)\,K(u',\ v')\,K^*(u'',\ v'')\times$$
$$\times\exp\left\{2\pi i\left[(f+f')\,u'+(g+g')\,v'-(f'+f'')\,u''-(g'+g'')\,v''\right]\right\}\times$$
$$\times df\,dg\,du'\,dv'\,du''\,dv''=\iint\limits_{-\infty}^{+\infty}\mathcal{Y}_0^-(f,g)\,\mathcal{K}(f+f',\,g+g')\,\mathcal{K}^*(f+f'',\,g+g'')\,df\,dg. \tag{47}$$

При переходе от четырехкратного интеграла к шестикратному мы подставили для J_0^- выражение из (45б), а при переходе от шестикратного к двукратному использовали преобразование, обратное (35).

Как мы видим, в (46) влияние предмета (характеризуемого функцией $\mathcal{T}$) и комбинированное влияние освещения ($\mathcal{Y}_0^-$) и оптической системы ($\mathcal{K}$) разделены. При равномерном освещении ($I_0^- = \text{const}$) интенсивность света, выходящего из объекта, пропорциональна $|F|^2$ и в случае совершенного отображения интенсивность в плоскости изображения выражается (с точностью до постоянного множителя) формулой

$$\tilde{I}_1(x_1, y_1) = F(x_1, y_1) F^*(x_1, y_1) = \iiiint\limits_{-\infty}^{+\infty} \mathcal{F}(f', g') \mathcal{F}^*(f'', g'') \times \\ \times \exp\{-2\pi i [(f'-f'') x_1 + (g'-g'') y_1]\} \, df' \, dg' \, df'' \, dg''. \qquad (48)$$

В соотношениях (46) и (48) истинная интенсивность I_1 и идеальная интенсивность $\tilde{I}_1$ выражены в виде суммы вкладов, вносимых всеми парами частот (f', g'), (f'', g'') пространственного спектра объекта. В первом случае каждый вклад в $\mathcal{T}$ раз больше, чем во втором. Отсюда следует, что если $\mathcal{T}$ не постоянно для всех значений f', g', f'', g'', для которых обе спектральные компоненты $\mathcal{F}(f', g')$ и $\mathcal{F}(f'', g'')$ отличны от нуля, то некоторая информация об объекте будет теряться или искажаться. Функция $\mathcal{T}$ называется *взаимным коэффициентом пропускания* системы, работающей при данном освещении проходящим светом.

Рассмотрим теперь вместо самой интенсивности ее пространственный спектр $\mathcal{I}(f, g)$. Для получения соответствующего выражения умножим обе части (46) на $\exp[2\pi i(fx_1 + gy_1)]$ и проинтегрируем по x_1 и y_1. Используя интегральную теорему Фурье (или, короче, представление дельта-функции Дирака в виде интеграла Фурье (см. приложение 4)), найдем

$$\mathcal{I}_1(f, g) = \iint\limits_{-\infty}^{+\infty} \mathcal{T}(f'+f, g'+g, f', g') \mathcal{F}(f'+f, g'+g) \mathcal{F}^*(f', g') \, df' \, dg'. \qquad (49)$$

Для идеального случая, которому соответствует выражение (48), получим

$$\tilde{\mathcal{I}}_1(f, g) = \iint\limits_{-\infty}^{+\infty} \mathcal{F}(f'+f, g'+g) \mathcal{F}^*(f', g') \, df' \, dg'. \qquad (50)$$

В этих формулах $\mathcal{I}_1$ и $\tilde{\mathcal{I}}_1$ выражены в виде сумм вкладов от каждой пространственной гармоники (f', g') структуры объекта. Как мы видим, $\mathcal{T}$ играет ту же роль, что и прежде. Эта величина характеризует изменения, возникающие в каждом вкладе и связанные со способом освещения объекта и характеристиками пропускания системы, формирующей изображение.

Так как функция отклика $\mathcal{K}(f, g)$ равна нулю, когда точка $\xi = \bar{\lambda} Rf$, $\eta = \bar{\lambda} Rg$ лежит вне выходного зрачка, то из (47) следует, что для достаточно высоких частот $\mathcal{T}$ обращается в нуль. Если выходным зрачком служит круг радиуса a, то произведение $\mathcal{K}(f+f', g+g') \mathcal{K}^*(f+f', g+g')$ и, следовательно, величина $\mathcal{T}(f', g'; f'', g'')$ могут отличаться от нуля только при условии, что в плоскости f, g отлична от нуля область наложения кругов C' и C'' с центрами в точках $O'(-f', -g')$, $O''(-f'', -g'')$ и одинаковыми радиусами $\sqrt{f^2+g^2} = a/\bar{\lambda} R$ (рис. 10.15). Чтобы проиллюстрировать влияние освещения, предположим, что применяется критическое освещение или освещение по Кёлеру и что числовая апертура $n'_c \sin\theta'_c$ конденсорной системы в m раз превышает числовую апертуру $n_0 \sin\theta_0$ системы, отображающей объект. Тогда, согласно (18) или (30), взаимная интенсивность освещающего пучка равна

$$J_0^-(x_0 - x'_0, y_0 - y'_0) = \left(\frac{2J_1(mv)}{mv}\right) I_0^-, \qquad (51)$$

где

$$v = \frac{2\pi}{\bar{\lambda}_0} \sqrt{(x_0' - x_0)^2 + (y_0' - y_0)^2}\, n_0 \sin \theta_0, \tag{52}$$

а I_0^- — интенсивность падающего света (она считается постоянной). В правой части выражения (51) стоит (см. п. 8.5.2) фурье-образ функции

$$\left.\begin{aligned} \mathscr{Y}_0^-(f, g) &= \text{const} = \left(\frac{\bar{\lambda}_0^2}{\pi m^2 n_0^2 \sin^2 \theta_0}\right) I_0^-, \quad \text{когда} \quad f^2 + g^2 < \frac{m^2 n_0^2 \sin^2 \theta_0}{\bar{\lambda}_0^2}, \\ \mathscr{Y}_0^-(f, g) &= 0, \qquad\qquad\qquad\qquad\qquad\quad \text{когда} \quad f^2 + g^2 > \frac{m^2 n_0^2 \sin^2 \theta_0}{\bar{\lambda}_0^2}. \end{aligned}\right\} \tag{53}$$

На рис. 10.15 круг, вне которого $\mathscr{Y}_0^-$ обращается в нуль, обозначен через C. Ясно, что для данных (f', g') и (f'', g'') вклад в интеграл (47) для $\mathscr{T}$ вносят лишь те точки плоскости f, g, которые лежат внутри области перекрытия кругов C, C' и C'' (на рис. 10.15 она заштрихована).

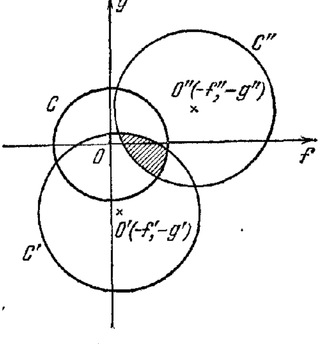


Рис. 10.15 Эффективная область интегрирования (заштрихована) для взаимного коэффициента пропускания $\mathscr{T}(f', g'; f'', g'')$ отображающей системы с круглым выходным зрачком радиуса a.

Предполагается, что объект освещается проходящим квазимонохроматическим светом со средней длиной волны $\bar{\lambda}_0$ через конденсор с числовой апертурой $mn_0 \sin \theta_0$, где $n_0 \sin \theta_0$—числовая апертура отображающей системы, C' и C''—круги радиуса $a/\bar{\lambda}R \approx n_0 \sin \theta_0 / M\bar{\lambda}_0$ с центрами в точках $O'(-f', -g')$ и $O''(-f'', -g'')$ соответственно, C—круг радиуса $mn_0 \sin \theta_0/\bar{\lambda}_0$ с центром в начале координат, R—радиус опорной сферы Гаусса, M – гауссово увеличение.

§ 10.6. Некоторые теоремы, касающиеся взаимной когерентности

В §§ 10.4 и 10.5 мы рассматривали интерференцию и дифракцию квазимонохроматического света и ограничивались случаем времени задержки τ, малом по сравнению с временем когерентности света. Мы показали, что при этом с хорошей точностью зависимость корреляционных функций от τ представлена только гармоническим членом, т. е. что

$$\Gamma(P_1, P_2, \tau) \approx J(P_1, P_2) \exp(-2\pi i \bar{\nu} \tau),$$
$$\gamma(P_1, P_2) \approx \mu(P_1, P_2) \exp(-2\pi i \bar{\nu} \tau).$$

Элементарная теория, использующая это приближение, позволяет учесть уменьшение видности в «центре» картины ($\tau = 0$), обусловленное конечными размерами источника света. Однако она не принимает во внимание изменения видности с увеличением разности хода. Для правильного описания явлений в случае, когда временем задержки τ нельзя пренебречь по сравнению с временем когерентности, необходимо использовать более точные выражения для корреляционных функций. Ниже мы рассмотрим соответствующее обобщение некоторых формул.

10.6.1. Расчет взаимной когерентности для света от некогерентного источника. Пусть $V_1(t)$ и $V_2(t)$— возмущения в точках P_1 и P_2 волнового поля, созданного протяженным (не обязательно квазимонохроматическим) первичным источником σ. Вначале будем считать, что среда между σ и точками P_1 и P_2 однородна.

Как и в п. 10.4.2, предположим, что источник разделен на элементы $d\sigma_1$, $d\sigma_2$, ..., линейные размеры которых малы по сравнению с эффективными длинами волн, а центры находятся близ точек S_1, S_2, ... Если $V_{m1}(t)$ и $V_{m2}(t)$ — вклады в V_1 и V_2, вносимые элементом $d\sigma_m$, то

$$V_1(t) = \sum_m V_{m1}(t), \qquad V_2(t) = \sum_m V_{m2}(t), \tag{1}$$

и взаимная функция когерентности определяется выражением

$$\Gamma(P_1, P_2, \tau) = \langle V_1(t+\tau) V_2^*(t)\rangle = \sum_m \langle V_{m1}(t+\tau) V_{m2}^*(t)\rangle. \qquad (2)$$

В правой части (2) мы пренебрегли членами типа $\langle V_{m1}(t+\tau) V_{n2}^*(t)\rangle$ $(m \neq n)$, так как можно полагать, что вклады, вносимые различными элементами источника, взаимно некогерентны.

Мы поступим далее несколько иначе, чем в § 10.4. Согласно (10.3.30) каждый член под знаком суммы в (2) можно представить в виде

$$\langle V_{m1}(t+\tau) V_{m2}^*(t)\rangle = 4\int_0^\infty G_m(P_1, P_2, \nu) \exp(-2\pi i\nu\tau)\, d\nu, \qquad (3)$$

где

$$G_m(P_1, P_2, \nu) = \lim_{T\to\infty}\left[\frac{v_{mT}(P_1, \nu)\, v_{mT}^*(P_2, \nu)}{2T}\right] \qquad (4)$$

— взаимная спектральная плотность возмущений $V_{m1} = V_m(P_1, t)$, $V_{m2} = V_m(P_2, t)$. Здесь V_m — вклад соответствующей частоты в возмущение, создаваемое элементом $d\sigma_m$. Этот вклад распространяется в виде сферической волны, поскольку предполагается, что среда однородна. Следовательно,

$$\left.\begin{aligned} v_{mT}(P_1, \nu) &= a_{mT}(\nu)\frac{\exp(ikR_{m1})}{R_{m1}}, \\ v_{mT}(P_2, \nu) &= a_{mT}(\nu)\frac{\exp(ikR_{m2})}{R_{m2}}, \end{aligned}\right\} \qquad (5)$$

где R_{m1} и R_{m2} — расстояния между точкой S_m источника и точками P_1 и P_2 (предполагается, что эти расстояния велики по сравнению с эффективными длинами волн), а $k = 2\pi\nu/v = 2\pi/\lambda$. Амплитуда $|a_m(\nu)|$ величины $a_m(\nu)$ определяет «силу» компоненты с частотой ν от элемента $d\sigma_m$, а $\arg a_m(\nu)$ — ее фазу. Из (4) и (5) следует, что

$$G_m(P_1, P_2, \nu) = \left\{\lim_{T\to\infty}\frac{|a_{mT}(\nu)|^2}{2T}\right\}\frac{\exp[ik(R_{m1} - R_{m2})]}{R_{m1}R_{m2}}. \qquad (6)$$

Величина, стоящая в фигурных скобках в правой части, является спектральной плотностью света, идущего от элемента $d\sigma_m$ источника. Предположим, как и в п. 10.4.2, что число элементов источника настолько велико, что его можно считать непрерывным. Следовательно, если $I(S_m, \nu)\, d\sigma_m\, d\nu = 4\lim_{T\to\infty}[|a_{mT}(\nu)|^2/2T]$, т. е. если $I(S, \nu)$ — интенсивность на единицу площади источника в единице частотного интервала, то, согласно (2), (3) и (6), имеем

$$\Gamma(P_1, P_2, \tau) = \sqrt{I(P_1)}\sqrt{I(P_2)}\,\gamma(P_1, P_2, \tau) = \int_0^\infty d\nu \exp(-2\pi i\nu\tau)\int_\sigma I(S, \nu)\frac{\exp[ik(R_1 - R_2)]}{R_1R_2}\, dS, \qquad (7)$$

где

$$\left.\begin{aligned} I(P_1) &= \Gamma(P_1, P_1, 0) = \int_0^\infty d\nu \int_\sigma \frac{I(S, \nu)}{R_1^2}\, dS, \\ I(P_2) &= \Gamma(P_2, P_2, 0) = \int_0^\infty d\nu \int_\sigma \frac{I(S, \nu)}{R_2^2}\, dS, \end{aligned}\right\} \qquad (8)$$

— интенсивности в точках P_1 и P_2, а R_1 и R_2 — расстояния от этих точек до точки S источника. Уравнения (7) служат обобщенной формой формул Ван-Циттерта — Цернике (10.4.20) и (10.4.21).

Если среда между источником и точками P_1 и P_2 неоднородна, мы можем поступить так же, как в § 10.4, а именно заменить множители $(\exp ikR_j)/R_j$ на $icK(S, P_j, \nu)/\nu$, где K — соответствующая функция пропускания среды. Тогда вместо (7) мы получим

$$\Gamma(P_1, P_2, \tau) = \sqrt{I(P_1)}\sqrt{I(P_2)}\,\gamma(P_1, P_2, \tau) = \\ = c^2\int_0^\infty \frac{d\nu}{\nu^2}\exp(-2\pi i\nu\tau)\int_\sigma I(S, \nu)K(S, P_1, \nu)K^*(S, P_2, \nu)\,dS, \tag{9}$$

где

$$\left.\begin{aligned} I(P_1) &= c^2\int_0^\infty \frac{d\nu}{\nu^2}\int_\sigma I(S, \nu)\,|K(S, P_1, \nu)|^2\,dS, \\ I(P_2) &= c^2\int_0^\infty \frac{d\nu}{\nu^2}\int_\sigma I(S, \nu)\,|K(S, P_2, \nu)|^2\,dS. \end{aligned}\right\} \tag{10}$$

По аналогии с преобразованиями, используемыми в § 10.4, перепишем (9) или (10) в несколько иной форме. Пусть

$$\left.\begin{aligned} \frac{ic}{\nu}K(S, P_1, \nu)\sqrt{I(S, \nu)} &= U(S, P_1, \nu), \\ \frac{ic}{\nu}K(S, P_2, \nu)\sqrt{I(S, \nu)} &= U(S, P_2, \nu). \end{aligned}\right\} \tag{11}$$

Тогда (9) и (10) примут вид

$$\Gamma(P_1, P_2, \tau) = \sqrt{I(P_1)}\sqrt{I(P_2)}\,\gamma(P_1, P_2, \tau) = \\ = \int_0^\infty d\nu\exp(-2\pi i\nu\tau)\int_\sigma U(S, P_1, \nu)U^*(S, P_2, \nu)\,dS, \tag{12}$$

где

$$I(P_1) = \int_0^\infty d\nu\int_\sigma |U(S, P_1, \nu)|^2\,dS, \qquad I(P_2) = \int_0^\infty d\nu\int_\sigma |U(S, P_2, \nu)|^2\,dS. \tag{13}$$

В формуле (12), которая является обобщением формулы Гопкинса (10.4.35), взаимная функция когерентности и комплексная степень когерентности выражены через распределение света, создаваемое ассоциированным воображаемым источником. В самом деле, согласно (11) величину $U(S, P, \nu)$ можно рассматривать как возмущение в точке P, обусловленное находящимся в точке S монохроматическим точечным источником с частотой ν, нулевой фазой и «силой», пропорциональной $\sqrt{I(S, \nu)}$.

10.6.2. Распространение взаимной когерентности. Как и в п. 10.4.4, вообразим, что поверхность $\mathcal{A}$ пересекает пучок света от протяженного первичного источника σ (см. рис. 10.8). Мы покажем, как, зная взаимную когерентность для всех пар точек на поверхности $\mathcal{A}$, можно определить ее значение на любой другой поверхности $\mathcal{B}$, освещаемой светом от $\mathcal{A}$. Для простоты предположим, что между $\mathcal{A}$ и $\mathcal{B}$ находится среда с показателем преломления, равным единице.

Взаимную когерентность для любой пары точек Q_1 и Q_2 на $\mathcal{B}$ можно рассчитать с помощью (12). В этой формуле U представляет собой монохроматическое возмущение, и следовательно, его можно определить на основе принципа Гюйгенса — Френеля при условии, что известны возмущения во всех точках поверхности $\mathcal{A}$; имеем

$$U(S, Q_1, \nu) = \int_{\mathcal{A}} U(S, P_1, \nu)\frac{\exp(ik_0 s_1)}{s_1}\Lambda_1\,dP_1. \tag{14}$$

Здесь s_1 означает расстояние между произвольной точкой P_1 на $\mathcal{A}$ и точкой Q_1, а Λ_1 — обычный коэффициент наклона. Совершенно аналогичное соотношение справедливо и для $U(S, Q_2, \nu)$. Подставляя эти формулы в (12) и изменяя порядок интегрирования, получим для взаимной когерентности следующее выражение:

$$\Gamma(Q_1, Q_2, \tau)= \\ =\iint\limits_{\mathcal{A}\mathcal{A}} \frac{dP_1\,dP_2}{s_1 s_2}\int\limits_0^\infty J(P_1, P_2, \nu)\,\Lambda_1(\nu)\,\Lambda_2^*(\nu)\exp\left\{-2\pi i\nu\left[\tau-\frac{s_1-s_2}{c}\right]\right\}d\nu, \quad (15)$$

где

$$J(P_1, P_2, \nu)=\int\limits_0 U(S, P_1, \nu)\,U^*(S, P_2, \nu)\,dS. \quad (16)$$

Двойное интегрирование в (15) означает, что точки P_1 и P_2 пробегают поверхность $\mathcal{A}$ независимо. Здесь коэффициенты наклона $\Lambda_1(\nu)$ и $\Lambda_2(\nu)$ зависят от частоты и медленно изменяются с ν по сравнению с изменением остальных членов. Если ширина эффективного спектрального интервала света достаточно мала, эти коэффициенты можно заменить на $\bar{\Lambda}_1=\Lambda_1(\bar{\nu})$, $\bar{\Lambda}_2=\Lambda_2(\bar{\nu})$, где $\bar{\nu}$ — средняя частота света. Согласно (12) интеграл по ν равен $\Gamma(P_1, P_2, \tau-(s_1-s_2)/c)$. Таким образом, окончательно получим

$$\Gamma(Q_1, Q_2, \tau)=\iint\limits_{\mathcal{A}\mathcal{A}} \frac{\Gamma\left(P_1, P_2, \tau-\frac{s_1-s_2}{c}\right)}{s_1 s_2}\,\bar{\Lambda}_1\bar{\Lambda}_2^*\,dP_1\,dP_2. \quad (17)$$

Это и есть искомая формула, в которой взаимная функция когерентности точек Q_1 и Q_2 поверхности $\mathcal{B}$ выражена через взаимную интенсивность всех пар точек поверхности $\mathcal{A}$.

Особый интерес представляет случай совпадения точек Q_1 и Q_2 и $\tau=0$ (общую точку обозначим через Q). Тогда левая часть соотношения (17) превращается в интенсивность $I(Q)$. Если, кроме того, мы подставим в правую часть функцию Γ, выраженную через интенсивности и коэффициент корреляции γ, то формула (17) примет вид

$$I(Q)=\iint\limits_{\mathcal{A}\mathcal{A}} \frac{\sqrt{I(P_1)}\,\sqrt{I(P_2)}}{s_1 s_2}\,\gamma\left(P_1, P_2, \frac{s_2-s_1}{c}\right)\bar{\Lambda}_1\bar{\Lambda}_2^*\,dP_1\,dP_2. \quad (18)$$

В этом соотношении *интенсивность* в произвольной точке Q выражена в виде суммы вкладов, вносимых всеми парами элементов произвольной поверхности $\mathcal{A}$. Каждый вклад пропорционален среднему геометрическому из интенсивностей в двух элементах, обратно пропорционален произведению расстояний этих элементов до Q и входит с весом, равным соответствующему значению коэффициента корреляции γ.

Соотношения (17) и (18), предложенные Вольфом, обобщают закон распространения, найденный Цернике (10.4.45), и формулу (10.4.46) для интенсивности в частично когерентном волновом поле.

§ 10.7. Строгая теория частичной когерентности *)

10.7.1. Волновые уравнения для взаимной когерентности. Некоторые теоремы, выведенные выше и относящиеся к корреляционным функциям, во многих чертах подобны теоремам, относящимся к самому комплексному возмущению. Например, формула Ван-Циттерта — Цернике (10.4.21) для комплексной

*) Материал, изложенный в этом параграфе, в основном базируется на исследованиях Вольфа [15, 71].

степени когерентности в плоскости, освещаемой протяженным квазимонохроматическим первичным источником, оказывается идентичной формуле для комплексного возмущения в дифракционной картине, возникающей при дифракции на отверстии, размеры и форма которого совпадают с размерами и формой источника. Другим примером служат законы распространения взаимной интенсивности (см. (10 4.45)); как мы видим, они похожи на принцип Гюйгенса — Френеля. Теоремы, относящиеся к комплексному возмущению, можно рассматривать как приближенные положения, вытекающие из некоторых строгих теорем, а именно формул Гельмгольца и Кирхгофа (см. (8.3.7), (8.3.13)). Последние вытекают из положения, согласно которому световое возмущение распространяется, как волна. Указанная аналогия наводит на мысль, что корреляция также распространяется, как волна и что наши теоремы являются приближенными формулировками каких-то соответствующих теорем типа Гельмгольца — Кирхгофа. Нетрудно показать, что это действительно так.

Рассмотрим стационарное волновое поле в вакууме. Пусть $V(P_1, t)$ и $V(P_2, t)$ — возмущения в точках P_1 и P_2 соответственно Удобно вначале выразить взаимную функцию когерентности в более симметричной форме

$$\Gamma(P_1, P_2, t_1, t_2) = \langle V(P_1, t_1+t) V^*(P_2, t_2+t)\rangle =$$
$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} V_T(P_1, t+t_1) V_T^*(P_2, t+t_2)\, dt. \tag{1}$$

Далее пусть

$$\nabla_1^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \tag{2}$$

— лапласиан по декартовым координатам точки P_1. Действуя этим оператором на соотношение (1) и изменяя порядок различных операций, получим

$$\nabla_1^2 \Gamma(P_1, P_2, t_1, t_2) = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} \{[\nabla_1^2 V_T(P_1, t+t_1)]\, V_T^*(P_2, t+t_2)\}\, dt. \tag{3}$$

Вещественная часть $V_T^{(r)}$ величины V_T представляет истинное физическое волновое поле (например, декартову компоненту электрической векторной волны) и, следовательно, удовлетворяет волновому уравнению

$$\nabla_1^2 V_T^{(r)}(P_1, t+t_1) = \frac{1}{c^2} \frac{\partial^2}{\partial t_1^2} V_T^{(r)}(P_1, t+t_1). \tag{4а}$$

Мнимая часть $V_T^{(i)}$ величины V_T, а значит и V_T также удовлетворяет волновому уравнению; последний результат получается сразу же, если применить преобразование Гильберта к обеим частям (4а) и использовать следующее утверждение: если каждая из двух функций получается одна из другой преобразованием Гильберта, то та же зависимость справедлива и для их производных. Следовательно *),

$$\nabla_1^2 V_T(P_1, t+t_1) = \frac{1}{c^2} \frac{\partial^2}{\partial t_1^2} V_T(P_1, t+t_1). \tag{4б}$$

Отсюда следует, что в правой части (3) ∇_1^2 можно заменить на $\partial^2/c^2\partial t_1^2$. Вновь

*) Здесь можно привести и другое доказательство. Так как $V_T^{(r)}$ удовлетворяет волновому уравнению, то каждая из его спектральных компонент $v_T(\nu)$ $(-\infty \leqslant \nu \leqslant \infty)$ удовлетворяет уравнению Гельмгольца. Но согласно уравнению (10.2.18б) спектр $V_T = V_T^{(r)} + iV_T^{(i)}$ равен $2v_T(\nu)$ при $\nu > 0$ и нулю при $\nu < 0$. Следовательно, каждая спектральная компонента функции V_T также будет удовлетворять уравнению Гельмгольца, а значит и сама функция V_T удовлетворяет волновому уравнению.

изменяя порядок операций, получим

$$\nabla_1^2\Gamma(P_1, P_2, t_1, t_2)=\frac{1}{c^2}\frac{\partial^2}{\partial t_1^2}\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty}V_T(P_1, t+t_1)V_T^*(P_2, t+t_2)\,dt,$$

т. е.

$$\nabla_1^2\Gamma(P_1, P_2, t_1, t_2)=\frac{1}{c^2}\frac{\partial^2\Gamma(P_1, P_2, t_1, t_2)}{\partial t_1^2}. \quad (5а)$$

Совершенно аналогичным образом находим

$$\nabla_2^2\Gamma(P_1, P_2, t_1, t_2)=\frac{1}{c^2}\frac{\partial^2\Gamma(P_1, P_2, t_1, t_2)}{\partial t_2^2}, \quad (5б)$$

где ∇_2^2 — лапласиан по координатам точки P_2.

Для стационарного поля Γ зависит от t_1 и t_2 лишь через разность $t_1-t_2=\tau$. Поэтому, как и раньше, можно записать $\Gamma(P_1, P_2, t_1, t_2)=\Gamma(P_1, P_2, \tau)$. В этом случае $\partial^2/\partial t_1^2=\partial^2/\partial t_2^2=\partial^2/d\tau^2$, и вместо (5) мы получим

$$\nabla_1^2\Gamma(P_1, P_2, \tau)=\frac{1}{c^2}\frac{\partial^2\Gamma(P_1, P_2, \tau)}{\partial\tau^2}, \quad (6а)$$

$$\nabla_2^2\Gamma(P_1, P_2, \tau)=\frac{1}{c^2}\frac{\partial\Gamma(P_1, P_2, \tau)}{\partial\tau^2}. \quad (6б)$$

Таким образом, *в вакууме взаимная функция когерентности удовлетворяет двум волновым уравнениям* *). Каждое из них описывает изменение взаимной когерентности, когда одна из точек (P_2 или P_1) фиксирована, а другая точка и параметр τ меняются. Величина τ представляет собой разность времен между моментами, в которые рассматривается корреляция в этих двух точках. Во всех экспериментах τ входит только в комбинации $c\tau=\Delta\mathcal{S}$, т. е. как разность хода. Таким образом, само время исключено из окончательного описания поля. Эта особенность теории частичной когерентности весьма привлекательна, так как в оптических волновых полях истинные временны́е изменения совершенно невозможно обнаружить. Основную величину в предложенной теории, взаимную функцию когерентности $\Gamma(P_1, P_2, \tau)$, можно непосредственно измерить, например, с помощью интерференционных экспериментов, описанных в §§ 10.3 и 10.4.

10.7.2. Строгая формулировка закона распространения взаимной когерентности. Обратимся вновь к стационарному волновому полю в вакууме. Пусть Q_1 и Q_2 — любые две точки поля и $\mathcal{A}$ — любая воображаемая поверхность, окружающая эти точки. Если ∇_1^2 — лапласиан по координатам точки Q_1, то, согласно (5а), мы получим

$$\nabla_1^2\Gamma(Q_1, Q_2, t_1, t_2)=\frac{1}{c^2}\frac{\partial^2\Gamma(Q_1, Q_2, t_1, t_2)}{\partial t_1^2}. \quad (7)$$

Отсюда следует, что мы вправе использовать для Γ интегральную формулу Кирхгофа (8.3.13). Таким образом, функцию $\Gamma(Q_1, Q_2, t_1, t_2)$ можно выразить через значения $[\Gamma(P_1, Q_2, t_1, t_2)]_1$, где P_1 принимает все положения на поверхности $\mathcal{A}$, а квадратные скобки с индексом 1 ($[\ldots]_1$) означают запаздывание относительно t_1, т. е.

$$[\Gamma(P_1, Q_2, t_1, t_2)]_1=\Gamma\left(P_1, Q_2, t_1-\frac{s_1}{c}, t_2\right). \quad (8)$$

---

*) Когда τ мало по сравнению с временем когерентности, то, согласно (10.4.10), $\Gamma(P_1, P_2, \tau)\approx J(P_1, P_2)\exp(-2\pi i\bar{\nu}\tau)$. Из (6) следует, что в рамках теории, оперирующей с квазимонохроматическими пучками, взаимная интенсивность J_{12} в вакууме с хорошей точностью подчиняется уравнениям Гельмгольца

$$\nabla_1^2 J(P_1, P_2)+\bar{k}^2 J(P_1, P_2)=0, \qquad \nabla_2^2 J(P_1, P_2)+\bar{k}^2 J(P_1, P_2)=0.$$

Здесь s_1— расстояние между P_1 и Q_1 (рис. 10.16). Выписывая формулу Кирхгофа в явном виде, получим

$$\Gamma(Q_1, Q_2, t_1, t_2)=\frac{1}{4\pi}\int\limits_{\mathcal{A}}\{f_1[\Gamma(P_1, Q_2, t_1, t_2)]_1+ +g_1\left[\frac{\partial}{\partial t_1}\Gamma(P_1, Q_2, t_1, t_2)\right]_1+h_1\left[\frac{\partial}{\partial n_1}\Gamma(P_1, Q_2, t_1, t_2)\right]_1\Big\}dP_1. \quad (9)$$

Здесь $\partial/\partial n_1$ означает дифференцирование вдоль внутренней нормали к $\mathcal{A}$ в точке P_1 и

$$f_1=\frac{\partial}{\partial n_1}\left(\frac{1}{s_1}\right), \quad g_1=-\frac{1}{cs_1}\frac{\partial s_1}{\partial n_1}, \quad h_1=-\frac{1}{s_1}. \quad (10)$$

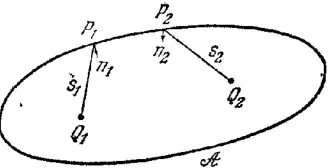


Рис. 10.16. Обозначения, используемые при строгой формулировке закона распространения взаимной когерентности.

Согласно (5б) мы получим также

$$\nabla_2^2\Gamma(P_1, Q_2, t_1, t_2)=\frac{1}{c^2}\frac{\partial^2\Gamma(P_1, Q_2, t_1, t_2)}{\partial t_2^2}, \quad (11)$$

где ∇_2^2— лапласиан по координатам точки Q_2. Следовательно, величину $\Gamma(P_1, Q_2, t_1, t_2)$, которая появляется в правой части (9), можно выразить в форме интеграла Кирхгофа, содержащего значения $[\Gamma(P_1, P_2, t_1, t_2)]_2$, где P_2 принимает все возможные положения на поверхности $\mathcal{A}$, а скобки с индексом 2 ($[\dots]_2$) означают запаздывание относительно t_2, т. е.

$$[\Gamma(P_1, P_2, t_1, t_2)]_2=\Gamma\left(P_1, P_2, t_1, t_2-\frac{s_2}{c}\right). \quad (12)$$

Здесь s_2 — расстояние между P_2 и Q_2. Соответствующая формула в явном виде запишется следующим образом:

$$\Gamma(P_1, Q_2, t_1, t_2)=\frac{1}{4\pi}\int\limits_{\mathcal{A}}\Big\{f_2[\Gamma(P_1, P_2, t_1, t_2)]_2+g_2\left[\frac{\partial}{\partial t_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2+ +h_2\left[\frac{\partial}{\partial n_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2\Big\}dP_2. \quad (13)$$

Здесь $\partial/\partial n_2$ означает дифференцирование вдоль внутренней нормали в точке P_2, а f_2, g_2, h_2 — те же величины, что и в (10), но с индексом 2. Продифференцировав (13) по t_1 и n_1, получим

$$\frac{\partial}{\partial t_1}\Gamma(P_1, Q_2, t_1, t_2)=\frac{1}{4\pi}\int\limits_{\mathcal{A}}\Big\{f_2\left[\frac{\partial}{\partial t_1}\Gamma(P_1, P_2, t_1, t_2)\right]_2+ +g_2\left[\frac{\partial^2}{\partial t_1\partial t_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2+h_2\left[\frac{\partial^2}{\partial t_1\partial n_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2\Big\}dP_2, \quad (14)$$

$$\frac{\partial}{\partial n_1}\Gamma(P_1, Q_2, t_1, t_2)=\frac{1}{4\pi}\int\limits_{\mathcal{A}}\Big\{f_2\left[\frac{\partial}{\partial n_1}\Gamma(P_1, P_2, t_1, t_2)\right]_2+ +g_2\left[\frac{\partial^2}{\partial n_1\partial t_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2+h_2\left[\frac{\partial^2}{\partial n_1\partial n_2}\Gamma(P_1, P_2, t_1, t_2)\right]_2\Big\}dP_2. \quad (15)$$

Подставляя (13), (14) и (15) в (9), найдем следующее выражение для $\Gamma(Q_1, Q_2, t_1, t_2)$:

$$\Gamma(Q_1, Q_2, t_1, t_2)=\frac{1}{(4\pi)^2}\int\limits_{\mathcal{A}}\int\limits_{\mathcal{A}}\Big\{f_1f_2[\Gamma]_{1,2}+f_1g_2\left[\frac{\partial}{\partial t_2}\Gamma\right]_{1,2}+ +f_1h_2\left[\frac{\partial}{\partial n_2}\Gamma\right]_{1,2}+g_1f_2\left[\frac{\partial}{\partial t_1}\Gamma\right]_{1,2}+g_1g_2\left[\frac{\partial^2}{\partial t_1\partial t_2}\Gamma\right]_{1,2}+g_1h_2\left[\frac{\partial^2}{\partial t_1\partial n_2}\Gamma\right]_{1,2}+ +h_1f_2\left[\frac{\partial}{\partial n_1}\Gamma\right]_{1,2}+h_1g_2\left[\frac{\partial^2}{\partial n_1\partial t_2}\Gamma\right]_{1,2}+h_1h_2\left[\frac{\partial^2}{\partial n_1\partial n_2}\Gamma\right]_{1,2}\Big\}dP_1\,dP_2. \quad (16)$$

Первыми двумя аргументами функции Г, стоящей в правой части (16), являются P_1 и P_2, а скобки $[\ldots]_{1,2}$ означают запаздывание относительно обеих переменных t, например,

$$[\Gamma]_{1,2}=\Gamma\left(P_1,\ P_2,\ t_1-\frac{s_1}{c},\quad t_2-\frac{s_2}{c}\right). \tag{17}$$

Наконец, используем предположение о стационарности, которое означает, что Г зависит лишь от разности временны́х аргументов. Запишем, как и раньше, $\Gamma(P_1,\ P_2,\ t_1,\ t_2)=\Gamma(P_1,\ P_2,\ \tau)$, $\tau=t_1-t_2$. Тогда $\partial/\partial t_1=-\partial/\partial t_2=\partial/\partial\tau$, и выражение (16) примет вид*)

$$\Gamma(Q_1, Q_2, \tau)=\frac{1}{(4\pi)^2}\iint\limits_{\mathcal{A}\mathcal{A}}\left\{f_1f_2[\Gamma]-f_1g_2\left[\frac{\partial}{\partial\tau}\Gamma\right]+f_1h_2\left[\frac{\partial}{\partial n_2}\Gamma\right]+\right.$$
$$+g_1f_2\left[\frac{\partial}{\partial\tau}\Gamma\right]-g_1g_2\left[\frac{\partial^2}{\partial\tau^2}\Gamma\right]+g_1h_2\left[\frac{\partial^2}{\partial\tau\,\partial n_2}\Gamma\right]+h_1f_2\left[\frac{\partial}{\partial n_1}\Gamma\right]-$$
$$\left.-h_1g_2\left[\frac{\partial^2}{\partial n_1\,\partial\tau}\Gamma\right]+h_1h_2\left[\frac{\partial^2}{\partial n_1\,\partial n_2}\Gamma\right]\right\}dP_1dP_2. \tag{18}$$

Первыми двумя аргументами функции Г, стоящей в правой части (18), являются P_1 и P_2, а скобки $[\ldots]$ означают «запаздывание» на величину $(s_1-s_2)/c$, например,

$$[\Gamma]=\Gamma\left(P_1,\ P_2,\ \tau-\frac{s_1-s_2}{c}\right). \tag{19}$$

Формулу (18) можно считать строгой формулировкой закона распространения взаимной когерентности (10.6.17). Она выражает значение взаимной функции когерентности для любых двух точек Q_1 и Q_2 через значения этой функции и некоторых ее производных для всех пар точек на произвольной замкнутой поверхности, окружающей обе эти точки.

В специальном случае совпадения точек Q_1 и Q_2 и $\tau=0$ мы получим из (18), подставляя $\Gamma_{12}(\tau)=\sqrt{I_1}\sqrt{I_2}\,\gamma_{12}(\tau)$, следующее выражение для интенсивности:

$$I(Q)=\frac{1}{(4\pi)^2}\iint\limits_{\mathcal{A}\mathcal{A}}\left(\sqrt{I_1}\sqrt{I_2}\left\{f_1f_2[\gamma]+(f_2g_1-f_1g_2)\left[\frac{\partial}{\partial\tau}\gamma\right]-g_1g_2\left[\frac{\partial^2}{\partial\tau^2}\gamma\right]\right\}+\right.$$
$$+\sqrt{I_1}\left\{f_1h_2\frac{\partial}{\partial n_2}(\sqrt{I_2}[\gamma])+g_1h_2\frac{\partial}{\partial n_2}\left(\sqrt{I_2}\left[\frac{\partial}{\partial\tau}\gamma\right]\right)\right\}+$$
$$+\sqrt{I_2}\left\{f_2h_1\frac{\partial}{\partial n_1}(\sqrt{I_1}[\gamma])-g_2h_1\frac{\partial}{\partial n_1}\left(\sqrt{I_1}\left[\frac{\partial}{\partial\tau}\gamma\right]\right)\right\}+$$
$$\left.+h_1h_2\frac{\partial^2}{\partial n_1\,\partial n_2}(\sqrt{I_1}\sqrt{I_2}[\gamma])\right)dP_1\,dP_2. \tag{20}$$

Здесь I_1 и I_2 — интенсивности в точках P_1 и P_2 соответственно, $[\gamma]=\gamma(P_1, P_2, (s_2-s_1)/c)$ и т. д. Формулу (20) можно считать строгой формулировкой теоремы, выражаемой уравнением (10.6.18). Она определяет интенсивность в произвольной точке Q через распределение интенсивности и комплексную степень когерентности (и некоторых производных от этих величин) на произвольной поверхности, окружающей Q.

10.7.3. Время когерентности и эффективная ширина спектра. Понятие времени когерентности, которое оказалось полезным при рассмотрении многих проблем, относящихся к полихроматическому свету, было введено в п. 7.5.8 при изучении возмущения, возникающего вследствие суперпозиции идентичных волновых цугов конечной длины. На простом примере (случайная последовательность периодических волновых цугов) мы показали, что время ко-

*) Уравнение (18) применимо к случаю распространения от замкнутой поверхности $\mathcal{A}$ произвольной формы. Для распространения от плоской поверхности формулы значительно упрощаются [72].

герентности *) $\Delta\tau$ и эффективная ширина спектра $\Delta\nu = c\Delta\lambda/\bar{\lambda}^2$ получающегося возмущения связаны по порядку величины соотношением

$$\Delta\tau\,\Delta\nu \approx 1. \qquad (21)$$

Мы упоминали также, что подобное соотношение выполняется и при более общих условиях, если под $\Delta\tau$ и $\Delta\nu$ понимать соответствующие средние величины. В настоящем разделе мы определим эти величины и строго установим искомое соотношение взаимности.

Предположим, что пучок света в точке P делится на два пучка, которые сводятся вновь после того, как между ними возникла разность хода $c\tau$. Получающиеся интерференционные эффекты характеризуются функцией автокогерентности

$$\Gamma(\tau) = \langle V(t+\tau)V^*(t)\rangle = 4\int_0^\infty G(\nu)\exp(-2\pi i\nu\tau)\,d\nu, \qquad (22)$$

где $V(t)$ — комплексное возмущение в точке P, а $G(\nu)$ — спектральная плотность.

Так как степень когерентности двух интерферирующих пучков выражается в виде $|\gamma(\tau)| = |\Gamma(\tau)|/\Gamma(0)$, разумно (и с математической точки зрения удобно) определить *время когерентности* $\Delta\tau$ света в точке P как нормированную среднеквадратичную «ширину» **) функции $|\Gamma(\tau)|^2$, т. е.***)

$$(\Delta\tau)^2 = \frac{\int_{-\infty}^{\infty}\tau^2|\Gamma(\tau)|^2\,d\tau}{\int_{-\infty}^{\infty}|\Gamma(\tau)|^2\,d\tau}. \qquad (23)$$

Определим далее *эффективную ширину спектра* $\Delta\nu$ света в точке P как нормированную среднеквадратичную ширину спектра Γ, т. е. как нормированную среднеквадратичную «ширину» квадрата спектральной плотности $G(\nu)$ в области $\nu \geqslant 0$. Таким образом,

$$(\Delta\nu)^2 = \frac{\int_0^\infty(\nu-\bar{\nu})^2G^2(\nu)\,d\nu}{\int_0^\infty G^2(\nu)\,d\nu}, \qquad \bar{\nu} = \frac{\int_0^\infty \nu G^2(\nu)\,d\nu}{\int_0^\infty G^2(\nu)\,d\nu}. \qquad (24)$$

Чтобы установить требуемое соотношение взаимности, положим

$$\left.\begin{array}{lr} \xi = \nu - \bar{\nu}, & \text{(а)}\\ \left.\begin{array}{ll}\Phi(\xi) = 4G(\bar{\nu}+\xi), & \text{когда } \xi > -\bar{\nu},\\ \Phi(\xi) = 0, & \text{когда } \xi < -\bar{\nu},\end{array}\right\} & \text{(б)}\\ \Psi(\tau) = \Gamma(\tau)\exp(2\pi i\bar{\nu}\tau). & \text{(в)}\end{array}\right\} \qquad (25)$$

---

*) В соответствии с обозначениями, принятыми в настоящей главе, вместо Δt мы здесь пишем $\Delta\tau$.

**) Здесь и ниже авторы определяют нормированную среднеквадратичную «ширину» функции через дисперсию аргумента при функции распределения, совпадающей с нормированным квадратом данной функции. (*Прим. ред.*)

***) Среднее значение $\bar{\tau} = \int_{-\infty}^{+\infty}\tau|\Gamma(\tau)|^2\,d\tau\Big/\int_{-\infty}^{+\infty}|\Gamma(\tau)|^2\,d\tau$ равно нулю, так как $|\Gamma(\tau)|$ — четная функция τ. Другое определение времени когерентности дано в работах [73].

Предположим, что функция $\Phi(\xi)$ непрерывна всюду $(-\infty<\xi<\infty)$; следовательно *), $\Phi(-\bar{\nu})=G(0)=0$. Из (22) вытекает, что Ψ и Φ служат фурье-образами друг друга, т. е.

$$\Psi(\tau)=\int_{-\infty}^{\infty}\Phi(\xi)\exp(-2\pi i\xi\tau)\,d\xi,\qquad \Phi(\xi)=\int_{-\infty}^{\infty}\Psi(\tau)\exp(2\pi i\xi\tau)\,d\tau. \tag{26}$$

Выражения для $\Delta\tau$ и $\Delta\nu$ принимают вид

$$(\Delta\tau)^2=\frac{1}{N}\int_{-\infty}^{\infty}\tau^2|\Psi(\tau)|^2\,d\tau, \tag{27}$$

$$(\Delta\nu)^2=\frac{1}{N}\int_{-\infty}^{\infty}\xi^2\Phi^2(\xi)\,d\xi, \tag{28}$$

где

$$N=\int_{-\infty}^{\infty}|\Psi(\tau)|^2\,d\tau=\int_{-\infty}^{\infty}\Phi^2(\xi)\,d\xi. \tag{29}$$

Выразим далее интеграл (28) через функцию Ψ. Используя второе из соотношений (26), получим

$$\begin{aligned}(\Delta\nu)^2&=\frac{1}{N}\int_{-\infty}^{\infty}\xi^2\Phi(\xi)\,d\xi\int_{-\infty}^{\infty}\Psi(\tau)\exp(2\pi i\xi\tau)\,d\tau=\\&=\frac{1}{N}\int_{-\infty}^{\infty}\Psi(\tau)\,d\tau\left(\frac{1}{2\pi i}\right)^2\frac{\partial^2}{\partial\tau^2}\int_{-\infty}^{\infty}\Phi(\xi)\exp(2\pi i\xi\tau)\,d\xi=\\&=-\frac{1}{4\pi^2}\frac{1}{N}\int_{-\infty}^{\infty}\Psi(\tau)\frac{\partial^2}{\partial\tau^2}\Psi^*(\tau)\,d\tau=\frac{1}{4\pi^2}\frac{1}{N}\int_{-\infty}^{\infty}\left|\frac{\partial\Psi}{\partial\tau}\right|^2d\tau.\end{aligned} \tag{30}$$

При переходе от второго выражения к третьему использовано первое из соотношений (26), а также соотношение $\Psi(-\tau)=\Psi^*(\tau)$. Последнее выражение получается из предыдущего при интегрировании по частям, если учесть, что $\Psi\to0$ при $\tau\to\pm\infty$; это справедливо, так как, по предположению, интеграл $\int_{-\infty}^{\infty}|\Psi(\tau)|^2\,d\tau$ в (29) сходится.

Из (27), (29) и (30) следует, что

$$(\Delta\tau)^2(\Delta\nu)^2=\frac{1}{16\pi^2}\left[\frac{4\left(\int_{-\infty}^{\infty}\tau^2|\Psi(\tau)|^2\,d\tau\right)\left(\int_{-\infty}^{\infty}\left|\frac{d\Psi}{d\tau}\right|^2d\tau\right)}{\left(\int_{-\infty}^{\infty}|\Psi(\tau)|^2\,d\tau\right)^2}\right]. \tag{31}$$

В приложении 8 с помощью простой алгебраической аргументации показано, что член в квадратных скобках в правой части (31) больше или равен единице для любой функции Ψ, для которой существуют интегралы. Таким образом, мы установили следующее *неравенство взаимности* **) для времени

*) Фактически это условие необходимо для сходимости интеграла в числителе (23). Если, как здесь предполагается, оно выполняется, то величина (23) не изменяется при замене $\Gamma(\tau)=\langle V(t+\tau)V^*(t)\rangle$ на вещественную корреляционную функцию $\Gamma^{(r)}(\tau)=\mathrm{Re}\,\Gamma(\tau)=2\langle V^{(r)}(t+\tau)V^{(r)}(t)\rangle$ (см. [71, 74]). Более общий случай $G(0)\neq0$ исследовался в работах [75, 76]. См. также [77].

**) Наш вывод аналогичен выводу соотношения неопределенности Гейзенберга, предложенному Вейлем и Паули [78].

когерентности и эффективной ширины спектра:

$$\Delta\tau\,\Delta\nu \geqslant \frac{1}{4\pi}. \qquad (32)$$

Напомним, что согласно (10.4.5) степень когерентности $|\gamma_{11}(\tau)| = |\Gamma_{11}(\tau)|/\Gamma_{11}(0)$ для двух интерферирующих пучков квазимонохроматического света с одинаковой интенсивностью равна видности $\mathscr{V}^{\circ}(\tau)$ полос в точке, соответствующей разности хода $c\tau$ между этими пучками. Следовательно, (23) можно переписать в виде

$$(\Delta\tau)^2 = \frac{\int\limits_{-\infty}^{\infty} \tau^2 \mathscr{V}^{\circ 2}(\tau)\,d\tau}{\int\limits_{-\infty}^{\infty} \mathscr{V}^{\circ 2}(\tau)\,d\tau} = \frac{\int\limits_{0}^{\infty} \tau^2 \mathscr{V}^{\circ 2}(\tau)\,d\tau}{\int\limits_{0}^{\infty} \mathscr{V}^{\circ 2}(\tau)\,d\tau}. \qquad (33)$$

Таким образом, *время когерентности $\Delta\tau$ для интерферирующих пучков с одинаковой интенсивностью равно нормированному среднеквадратичному значению «ширины» квадрата функции видности.*

Данное определение времени когерентности является более удовлетворительным, чем определение, приведенное в п. 7.5.8, поскольку мы здесь не делали специальных предположений о природе элементарных полей, вызывающих возмущение. В самом деле, мы больше не требуем знания деталей поведения быстро флуктуирующей функции $V(t)$, а наше определение основывается на поддающейся измерению корреляционной функции $\Gamma(\tau)$. Если бы мы хотели сохранить описание интерференционных явлений с помощью элементарных волновых цугов, нам нужно было бы рассматривать $\Delta\tau$ как длительность *среднего* волнового цуга; однако такую интерпретацию следует применять с осторожностью.

Возвращаясь к соотношению (32), мы видим, что знак равенства получается лишь тогда, когда член в квадратных скобках в правой части (31) равен единице; согласно приложению 8 это возможно только в том случае, если $\Psi(\tau)$ — функция Гаусса. Но так как фурье-образ функции Гаусса тоже является функцией Гаусса, а эта последняя отлична от нуля при всех значениях ее аргумента ($-\infty < \xi < \infty$), то она не удовлетворяет второму условию (25б). Таким образом, знак равенства в (32) никогда не достигается. Однако если частота, соответствующая максимуму функции Гаусса, велика по сравнению со среднеквадратичным значением «ширины» этой функции, то вкладом в $\bar{\nu}$ и $\Delta\nu$, обусловленным отрицательным частотным интервалом, можно пренебречь, и очевидно, что для высокочастотного спектра, встречающегося в оптике, величина произведения $\Delta\tau\,\Delta\nu$ не может заметно отличаться от значения, которое соответствует всей кривой Гаусса. Таким образом, знак неравенства в (32) можно заменить знаком порядка величины, т. е.

$$\Delta\tau\,\Delta\nu \sim \frac{1}{4\pi}. \qquad (34)$$

Приведенное выше определение времени когерентности пригодно, если два интерферирующих пучка получаются из одного делением в точке P. Однако его легко распространить на случаи, когда два интерферирующих пучка образуются делением в двух точках P_1 и P_2, как, например, в интерференционном эксперименте Юнга. При этом вместо функции автокогерентности $\Gamma(\tau) = \Gamma(P, P, \tau)$ следует использовать взаимную функцию когерентности $\Gamma_{12}(\tau) = \Gamma(P_1, P_2, \tau)$, а вместо обычной спектральной плотности $G(\nu)$ — взаимную спектральную плотность $G_{12}(\nu)$. Единственное различие возникает из-за того, что величина $G_{12}(\nu)$ комплексна, а $\Gamma_{12}(\tau)$ — не обязательно четная функция τ

и, следовательно, $\bar{\tau}$ не обязательно равно нулю. Соответствующие выражения имеют вид

$$(\Delta\tau_{12})^2 = \frac{\int\limits_{-\infty}^{\infty} (\tau - \bar{\tau}_{12})^2 \, |\Gamma_{12}(\tau)|^2 \, d\tau}{\int\limits_{-\infty}^{\infty} |\Gamma_{12}(\tau)|^2 \, d\tau}, \quad \bar{\tau}_{12} = \frac{\int\limits_{-\infty}^{\infty} \tau \, |\Gamma_{12}(\tau)|^2 \, d\tau}{\int\limits_{-\infty}^{\infty} |\Gamma_{12}(\tau)|^2 \, d\tau}, \tag{35}$$

$$(\Delta\nu_{12})^2 = \frac{\int\limits_{0}^{\infty} (\nu - \bar{\nu}_{12})^2 \, |G_{12}(\nu)|^2 \, d\nu}{\int\limits_{0}^{\infty} |G_{12}(\nu)|^2 \, d\nu}, \quad \bar{\nu}_{12} = \frac{\int\limits_{0}^{\infty} \nu \, |G_{12}(\nu)|^2 \, d\nu}{\int\limits_{0}^{\infty} |G_{12}(\nu)|^2 \, d\nu}. \tag{36}$$

Величину $\Delta\tau_{12}$ можно назвать *взаимным временем когерентности*, а $\Delta\nu_{12}$ — *взаимной эффективной шириной спектра* светового возмущения в точках P_1 и P_2. Изменяя очевидным образом аргументацию, приводившуюся для вывода (32), легко показать, что эти величины удовлетворяют неравенству взаимности

$$(\Delta\tau_{12})(\Delta\nu_{12}) > 1/4\pi. \tag{37}$$

Наконец, обобщая соотношение (33), для квазимонохроматического света получим

$$(\Delta\tau_{12})^2 = \frac{\int\limits_{-\infty}^{\infty} (\tau - \bar{\tau}_{12})^2 \, \mathscr{V}^2_{12}(\tau) \, d\tau}{\int\limits_{-\infty}^{\infty} \mathscr{V}^2_{12}(\tau) \, d\tau}, \quad \bar{\tau}_{12} = \frac{\int\limits_{-\infty}^{\infty} \tau \mathscr{V}^2_{12}(\tau) \, d\tau}{\int\limits_{-\infty}^{\infty} \mathscr{V}^2_{12}(\tau) \, d\tau}. \tag{38}$$

Здесь $\mathscr{V}^2_{12}(\tau)$ — видность полос, образованных интерферирующими световыми пучками одинаковой интенсивности, идущими от точек P_1 и P_2.

§ 10.8. Поляризация квазимонохроматического света

В предыдущих разделах настоящей главы мы рассматривали световое возмущение как скалярную величину. Ниже мы кратко рассмотрим некоторые векторные свойства квазимонохроматической световой волны.

В § 1.4 мы указывали, что строго монохроматический свет всегда *поляризован*, т. е. что конец электрического (а также магнитного) вектора в каждой точке пространства движется периодически, описывая эллипс, который, конечно, в особых случаях переходит в круг или прямую линию. Мы рассматривали также *неполяризованный свет*. В этом случае можно считать, что конец вектора движется совершенно нерегулярно, и такие световые колебания не имеют никаких преимущественных направлений в плоскости, перпендикулярной к направлению распространения. Подобно случаям полной когерентности и полной некогерентности, указанные два случая также относятся к экстремальным. В общем случае изменение векторов поля не является ни вполне регулярным, ни вполне нерегулярным, и можно сказать, что свет *частично поляризован*. Обычно такой свет получается из неполяризованного при отражении (см. п. 1.5.3) или рассеянии (см. п. 13.5.2). Здесь мы исследуем основные свойства частично поляризованной световой волны Мы покажем, что для такой волны все наблюдаемые явления зависят от интенсивностей двух произвольных взаимно ортогональных компонент электрического вектора, перпендикулярных к направлению распространения, и от существующей между ними корреляции.

10.8.1. Матрица когерентности квазимонохроматической плоской волны *). Рассмотрим квазимонохроматическую световую волну со средней частотой $\bar{\nu}$, распространяющуюся в положительном направлении оси z. Пусть

$$E_x(t)=a_1(t)\exp\{i[\varphi_1(t)-2\pi\bar{\nu}t]\},\qquad E_y(t)=a_2(t)\exp\{i[\varphi_2(t)-2\pi\bar{\nu}t]\} \tag{1}$$

— две взаимно ортогональные компоненты электрического вектора в точке O, перпендикулярные к направлению распространения. Вновь воспользуемся комплексным представлением, рассмотренным в § 10.2; при этом E_x и E_y являются «аналитическими сигналами», ассоциированными с истинными (вещественными) компонентами $E_x^{(r)}=a_1(t)\cos[\varphi_1(t)-2\pi\bar{\nu}t]$, $E_y^{(r)}=a_2(t)\times\cos[\varphi_2(t)-2\pi\bar{\nu}t]$. Если бы свет был строго монохроматическим, то a_1, a_2, φ_1 и φ_2 были бы постоянными. Для квазимонохроматической волны эти величины зависят также от времени t, но, как мы видели, за любой интервал времени, малый по сравнению с временем когерентности, т. е. малый по сравнению с величиной, обратной эффективной спектральной ширине света $\Delta\nu$, их изменение относительно невелико.

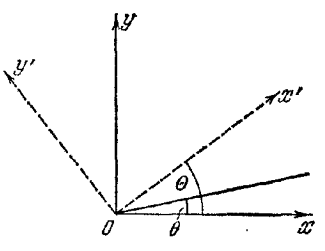


Рис. 10.17. Обозначения, используемые при вычислении матриц когерентности.

Предположим, что запаздывание y-компоненты электрического вектора относительно x-компоненты равно ε (это можно осуществить, например, с помощью одного из компенсаторов, описанных в п. 14.4.2), и рассмотрим интенсивность $I(\theta, \varepsilon)$ световых колебаний в направлении, которое образует угол θ с положительным направлением оси x (рис. 10.17). Такие колебания можно выделить, пропуская свет через соответствующим образом ориентированный поляризатор (см. п. 14.4.1).

Компонента электрического вектора в указанном направлении после введения запаздывания ε запишется в виде

$$E(t;\ \theta,\ \varepsilon)=E_x\cos\theta+E_y e^{i\varepsilon}\sin\theta, \tag{2}$$

так что

$$I(\theta,\ \varepsilon)=\langle E(t;\ \theta,\ \varepsilon)\,E^*(t;\ \theta,\ \varepsilon)\rangle=$$
$$=J_{xx}\cos^2\theta+J_{yy}\sin^2\theta+J_{xy}e^{-i\varepsilon}\cos\theta\sin\theta+J_{yx}e^{i\varepsilon}\sin\theta\cos\theta, \tag{3}$$

где J_{xx}, ... — элементы матрицы

$$\mathsf{J}=\begin{bmatrix}\langle E_xE_x^*\rangle & \langle E_xE_y^*\rangle\\ \langle E_yE_x^*\rangle & \langle E_yE_y^*\rangle\end{bmatrix}=\begin{bmatrix}\langle a_1^2\rangle & \langle a_1a_2\exp[i(\varphi_1-\varphi_2)]\rangle\\ \langle a_1a_2\exp[-i(\varphi_1-\varphi_2)]\rangle & \langle a_2^2\rangle\end{bmatrix}. \tag{4}$$

Диагональные элементы матрицы J вещественны и, как мы видим, представляют собой интенсивности x- и y-компонент электрического вектора. Следовательно, шпур Sp J нашей матрицы (т. е. сумма ее диагональных элементов) равен полной интенсивности света

$$\mathrm{Sp}\,\mathsf{J}=J_{xx}+J_{yy}=\langle E_xE_x^*\rangle+\langle E_yE_y^*\rangle. \tag{5}$$

Недиагональные элементы в общем случае комплексны, но они являются сопряженными. (Такая матрица, для которой $J_{ji}=J_{ij}^*$ при всех i и j, называется *эрмитовой*.)

Как и раньше (см. (10.4.9б)), нормируем смешанный член J_{xy}, полагая

$$\mu_{xy}=|\mu_{xy}|\exp(i\beta_{xy})=\frac{J_{xy}}{\sqrt{J_{xx}}\sqrt{J_{yy}}}. \tag{6}$$

*) Результаты, изложенные в пп. 10.8.1 и 10.8.2, основаны на исследованиях Вольфа [33]. Дальнейшее их развитие изложено в работе [79].

Тогда из неравенства Шварца тем же способом, что и при выводе (10.3.17), получим

$$|\mu_{xy}| \leqslant 1. \tag{7}$$

Комплексный коэффициент корреляции μ_{xy} играет примерно ту же роль, что и комплексная степень когерентности μ_{12}, введенная в п. 10.4.1. Он служит мерой корреляции между x- и y-компонентами электрического вектора. Модуль $|\mu_{xy}|$ служит мерой их «степени когерентности», а фаза β_{xy} этого коэффициента — мерой их «эффективной разности фаз». Матрица J называется *матрицей когерентности* световой волны. Так как J_{xx} и J_{yy} не могут быть отрицательными, (6) и (7) означают, что определитель матрицы когерентности неотрицателен, т. е. что

$$|\mathrm{J}| = J_{xx}J_{yy} - J_{xy}J_{yx} \geqslant 0. \tag{8}$$

Если использовать соотношение $J_{yx} = J^*_{xy}$ и обозначить символом Re вещественную часть, то (3) примет вид

$$I(\theta,\ \varepsilon) = J_{xx}\cos^2\theta + J_{yy}\sin^2\theta + 2\cos\theta\sin\theta\,\mathrm{Re}\,(J_{xy}\exp(-i\varepsilon)) =$$
$$= J_{xx}\cos^2\theta + J_{yy}\sin^2\theta + 2\sqrt{J_{xx}}\sqrt{J_{yy}}\cos\theta\sin\theta\,|\mu_{xy}|\cos[\beta_{xy} - \varepsilon], \tag{9}$$

при переходе от первой ко второй строке сделана подстановка (6). Если положить $J_{xx}\cos^2\theta = I^{(1)}$, $J_{yy}\sin^2\theta = I^{(2)}$, то последняя формула становится идентичной основному закону интерференции (10.4.11) для квазимонохроматических волновых полей.

Как и функции когерентности, которые мы рассматривали ранее, элементы матрицы когерентности заданной волны можно определить посредством относительно простых экспериментов. Это можно сделать различными способами. Необходимо лишь измерить интенсивность для нескольких различных значений θ (ориентации поляризатора) и ε (запаздывания, обусловленного компенсатором) и решить соответствующие соотношения, полученные из (3). Пусть $\{\theta,\ \varepsilon\}$ обозначает результаты измерений, соответствующие определенной паре значений θ, ε. Удобно использовать следующие их значения:

$$\{0°,\ 0\},\ \{45°,\ 0\},\ \{90°,\ 0\},\ \{135°,\ 0\},\ \left\{45°,\ \frac{\pi}{2}\right\},\ \left\{135°,\ \frac{\pi}{2}\right\}. \tag{10}$$

Из (3) вытекает, что элементы матрицы когерентности выражаются через интенсивности, полученные в результате измерений при шести указанных значениях в виде

$$\left.\begin{aligned} J_{xx} &= I(0°,\ 0), \qquad J_{yy} = I(90°,\ 0),\\ J_{xy} &= \frac{1}{2}\{I(45°,\ 0) - I(135°,\ 0)\} + \frac{1}{2}i\left\{I\left(45°, \frac{\pi}{2}\right) - I\left(135°, \frac{\pi}{2}\right)\right\},\\ J_{yx} &= \frac{1}{2}\{I(45°,\ 0) - I(135°,\ 0)\} - \frac{1}{2}i\left\{I\left(45°, \frac{\pi}{2}\right) - I\left(135°, \frac{\pi}{2}\right)\right\}. \end{aligned}\right\} \tag{11}$$

Как мы видим, для определения J_{xx}, J_{yy} и вещественной части J_{xy} (или J_{yx}) необходим лишь поляризатор. Величины J_{xx} и J_{yy} можно определить из измерений с поляризатором, ориентированным так, чтобы пропускать компоненты с азимутами $\theta = 0°$ и $\theta = 90°$ соответственно. Для получения вещественной части J_{xy} необходимы измерения с поляризатором, вначале ориентированным так, что он пропускает компоненту с азимутом $\theta = 45°$, а затем — компоненту с азимутом $\theta = 135°$. Для определения мнимой части J_{xy} (или J_{yx}) требуется также, согласно двум последним соотношениям в (11), компенсатор, который вносил бы разности фаз в четверть периода между x- и y-компонентами (например, четвертьволновая пластинка, см. п. 14.4.2). Поляризатор при этом вначале ориентирован так, что он пропускает компоненту с азимутом $\theta = 45°$, а затем — компоненту с азимутом $\theta = 135°$. В п. 14.4.2 мы покажем, что последние два измерения нужны для идентификации правой и левой круговой поляризации.

Из выражения (9) безусловно следует, что два пучка света с одинаковыми матрицами когерентности эквивалентны в том смысле, что в ряде аналогичных экспериментов с поляризатором и компенсатором получаются одинаковые (усредненные по времени) интенсивности *).

Посмотрим теперь, как меняется наблюдаемая интенсивность $I(0, \varepsilon)$ данной волны, когда один из аргументов (θ или ε) фиксирован, а другой изменяется. Предположим вначале, что мы фиксируем θ и изменяем ε. Из (9) следует, что интенсивность при этом будет меняться синусоидально между значениями

$$\left.\begin{aligned}(I(\varepsilon))_{\text{макс}} &= J_{xx}\cos^2\theta + J_{yy}\sin^2\theta + 2\,|J_{xy}|\sin\theta\cos\theta,\\ (I(\varepsilon))_{\text{мин}} &= J_{xx}\cos^2\theta + J_{yy}\sin^2\theta - 2\,|J_{xy}|\sin\theta\cos\theta.\end{aligned}\right\} \qquad (12)$$

Следовательно,

$$\frac{(I(\varepsilon))_{\text{макс}} - (I(\varepsilon))_{\text{мин}}}{(I(\varepsilon))_{\text{макс}} + (I(\varepsilon))_{\text{мин}}} = \frac{|J_{xy}|\sin 2\theta}{J_{xx}\cos^2\theta + J_{yy}\sin^2\theta}. \qquad (13)$$

Уравнение (13) открывает другой путь определения абсолютного значения J_{xy} (а следовательно, и $|\mu_{xy}|$) Оно показывает, что эту величину можно получить, измеряя J_{xx}, J_{yy}, $(I(\varepsilon))_{\text{макс}}$ и $(I(\varepsilon))_{\text{мин}}$; фазу величины J_{xy} легко найти, измеряя значения ε, при которых наблюдаются максимумы и минимумы интенсивности. Так, согласно (9),

$$\left.\begin{aligned}I &= (I(\varepsilon))_{\text{макс}}, \quad \text{когда}\quad \varepsilon = \beta_{xy} \pm 2m\pi \qquad (m = 0, 1, 2, \ldots),\\ I &= (I(\varepsilon))_{\text{мин}}, \quad \text{когда}\quad \varepsilon = \beta_{xy} \pm (2m+1)\pi \quad (m = 0, 1, 2, \ldots).\end{aligned}\right\} \qquad (14)$$

Чтобы выяснить, как изменяется интенсивность при фиксированном ε и переменном θ, удобно переписать (9) в несколько иной форме. Легко показать, что

$$I(\theta, \varepsilon) = \frac{1}{2}(J_{xx} + J_{yy}) + R\cos(2\theta - \alpha), \qquad (15)$$

где

$$R = \frac{1}{2}\sqrt{(J_{xx} - J_{yy})^2 + 4J_{xy}J_{yx}\cos^2(\beta_{xy} - \varepsilon)},$$
$$\operatorname{tg}\alpha = \frac{2\,|J_{xy}|\cos(\beta_{xy} - \varepsilon)}{J_{xx} - J_{yy}}. \qquad (16)$$

Из (15) видно, что при изменении θ интенсивность также меняется синусоидально. Ее экстремумы равны

$$\left.\begin{aligned}(I(\theta))_{\text{макс}} &= \frac{1}{2}(J_{xx} + J_{yy}) + R,\\ (I(\theta))_{\text{мин}} &= \frac{1}{2}(J_{xx} + J_{yy}) - R.\end{aligned}\right\} \qquad (17)$$

В правой части (17) только величина R зависит от ε. Она достигает максимального значения, когда $\cos^2(\beta_{xy} - \varepsilon) = 1$, т. е когда ε принимает одно из значений, указанных в (14). При этом величина R равна

$$(R(\varepsilon))_{\text{макс}} = \frac{1}{2}\sqrt{(J_{xx} - J_{yy})^2 + 4J_{xy}J_{yx}} = \frac{1}{2}(J_{xx} + J_{yy})\sqrt{1 - \frac{4\,|J|}{(J_{xx} + J_{yy})^2}}, \qquad (18)$$

где $|J|$ — детерминант (8) матрицы когерентности. Отсюда следует, что абсо-

*) Это утверждение справедливо только в рамках приближения квазимонохроматической теории, поскольку лишь в области ее применимости справедливо выражение (9) для интенсивности Поведение двух пучков может оказаться совершенно различным, если фазовой задержкой ε двух перпендикулярных друг другу компонент нельзя пренебречь по сравнению с длиной когерентности, измеренной в единицах средней длины волны $\bar{\lambda}$. Для более полного описания наблюдаемых свойств пучка необходимо вводить более общие матрицы когерентности, характеризующие корреляцию между компонентами в различные моменты времени, а также в различных точках (см [16], а также [32, 80, 81]).

лютные максимум и минимум интенсивности (при изменении и θ и ε) запишутся в виде

$$\left.\begin{aligned}(I(\theta,\ \varepsilon))_{\text{макс}} &= \frac{1}{2}(J_{xx}+J_{yy})\left[1+\sqrt{1-\frac{4|J|}{(J_{xx}+J_{yy})^2}}\right],\\ (I(\theta,\ \varepsilon))_{\text{мин}} &= \frac{1}{2}(J_{xx}+J_{yy})\left[1-\sqrt{1-\frac{4|J|}{(J_{xx}+J_{yy})^2}}\right].\end{aligned}\right\} \tag{19}$$

Следовательно,

$$\frac{(I(\theta,\ \varepsilon))_{\text{макс}}-(I(\theta,\ \varepsilon))_{\text{мин}}}{(I(\theta,\ \varepsilon))_{\text{макс}}+(I(\theta,\ \varepsilon))_{\text{мин}}}=\sqrt{1-\frac{4|J|}{(J_{xx}+J_{yy})^2}}. \tag{20}$$

Позже мы увидим, что эта величина имеет простой физический смысл.

До сих пор мы относили колебания электрического вектора к произвольной, но фиксированной прямоугольной системе координат xOy. Ниже мы рассмотрим, как преобразуется матрица когерентности при изменении положения осей. Пусть в плоскости, перпендикулярной к направлению распространения, новая прямоугольная система координат $x'Oy'$ выбрана так, что ось Ox' образует угол Θ с Ox (см. рис. 10.17). В новой системе координат компоненты электрического вектора выражаются через E_x, E_y следующим образом:

$$\left.\begin{aligned}E_{x'} &= E_x\cos\Theta+E_y\sin\Theta,\\ E_{y'} &= -E_x\sin\Theta+E_y\cos\Theta.\end{aligned}\right\} \tag{21}$$

Элементами трансформированной матрицы когерентности J' служат

$$J_{k'l'}=\langle E_{k'}E_{l'}^{*}\rangle, \tag{22}$$

где каждый из индексов k' и l' принимает значения x' и y'. Из (21) и (22) следует, что

$$J'=\begin{bmatrix}J_{xx}c^2+J_{yy}s^2+(J_{xy}+J_{yx})cs & (J_{yy}-J_{xx})cs+J_{xy}c^2-J_{yx}s^2\\ (J_{yy}-J_{xx})cs+J_{yx}c^2-J_{xy}s^2 & J_{xx}s^2+J_{yy}c^2-(J_{xy}+J_{yx})cs\end{bmatrix}, \tag{23}$$

где

$$c=\cos\Theta,\qquad s=\sin\Theta. \tag{24}$$

Как мы видим, шпур этой матрицы инвариантен относительно вращения системы координат. Легко показать, что ее детерминант также инвариантен относительно такого преобразования. Оба эти результата следуют и из хорошо известных теорем матричной алгебры.

Рассмотрим далее форму матрицы когерентности для некоторых случаев, представляющих особый интерес.

а. Неполяризованный (естественный) свет. Свет, с которым мы чаще всего сталкиваемся в природе, обладает тем свойством, что интенсивность любой его компоненты, перпендикулярной к направлению распространения, одинакова. Более того, на эту интенсивность не оказывает влияния никакое предыдущее взаимное запаздывание перпендикулярных друг другу компонент, на которые можно разложить свет. Другими словами,

$$I(\theta,\ \varepsilon)=\text{const} \tag{25}$$

для всех значений θ и ε. Такой свет называется *полностью неполяризованным*; часто его называют *естественным светом*.

Из (9) очевидно, что $I(\theta,\ \varepsilon)$ не зависит от ε и θ тогда и только тогда, когда

$$\mu_{xy}=0\quad\text{и}\quad J_{xx}=J_{yy}. \tag{26а}$$

Первое условие означает, что E_x и E_y взаимно некогерентны. На основании (6) и соотношения $J_{yx}=J_{xy}^{*}$ условия (26а) можно также записать в виде

$$J_{xy}=J_{yx}=0,\qquad J_{xx}=J_{yy}. \tag{26б}$$

Отсюда следует, что матрица когерентности для естественного света с

интенсивностью $J_{xx}+J_{yy}=I$ равна

$$\frac{1}{2}I\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}. \tag{27}$$

б. Полностью поляризованный свет. Рассмотрим вначале случай строго монохроматического света. Тогда a_1, a_2, φ_1 и φ_2 в (1) не зависят от времени, и матрица когерентности имеет вид

$$\begin{bmatrix}a_1^2 & a_1a_2e^{i\delta}\\ a_1a_2e^{-i\delta} & a_2^2\end{bmatrix}, \tag{28}$$

где *)

$$\delta=\varphi_1-\varphi_2. \tag{29}$$

Мы видим, что в этом случае

$$|\mathsf{J}|=J_{xx}J_{yy}-J_{xy}J_{yx}=0, \tag{30}$$

т. е. детерминант матрицы когерентности равен нулю. Тогда для комплексной степени когерентности компонент E_x и E_y имеем

$$\mu_{xy}=\frac{J_{xy}}{\sqrt{J_{xx}}\sqrt{J_{yy}}}=e^{i\delta}, \tag{31}$$

т. е. ее абсолютное значение равно единице (полная когерентность), а ее фаза равна разности фаз обеих компонент.

В специальном случае *линейно* поляризованного света (см. (1.4.33)) $\delta=m\pi$ ($m=0,\ \pm1,\ \pm2,\ \ldots$). Следовательно, матрица когерентности для линейно поляризованного света имеет вид

$$\begin{bmatrix}a_1^2 & (-1)^m a_1a_2\\ (-1)^m a_1a_2 & a_2^2\end{bmatrix}. \tag{32}$$

Электрический вектор колеблется в направлении, задаваемом соотношением $E_y/E_x=(-1)^m a_2/a_1$. В частности, каждая из матриц

$$I\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\qquad I\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} \tag{33}$$

соответствует линейно поляризованному свету интенсивности I с электрическим вектором, направленным по оси x ($a_2=0$) и оси y ($a_1=0$) соответственно. Матрицы

$$\frac{1}{2}I\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\qquad \frac{1}{2}I\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{34}$$

соответствуют линейно поляризованному свету интенсивности I с электрическим вектором, направленным соответственно под углами 45° и 135° к оси x ($a_1=a_2$, $m=0$ и $a_1=a_2$, $m=1$).

Для света, поляризованного по кругу (см. (1.4.35) и (1.4.36)), $a_1=a_2$, $\delta=m\pi/2$ ($m=\pm1,\ \pm3,\ \ldots$), и, значит, матрица когерентности имеет вид

$$\frac{1}{2}I\begin{bmatrix}1 & \pm i\\ \mp i & 1\end{bmatrix}, \tag{35}$$

где I — интенсивность света. На основании (1.4.38) и (1.4.40) верхний или нижний знак соответствует правой или левой поляризации.

Условие (30) может выполняться и для немонохроматического света, если зависимость величин a_1, a_2, φ_1 и φ_2 от времени такова, что *отношение амплитуд и разность фаз* не зависят от времени, т. е.

$$\frac{a_2(t)}{a_1(t)}=q,\qquad \delta=\varphi_1(t)-\varphi_2(t)=\chi, \tag{36}$$

*) Чтобы в дальнейшем использовать некоторые результаты, изложенные в § 1.4, заметим, что φ_1 и φ_2 соответствуют $-\delta_1$ и $-\delta_2$, введенным в п. 1.4.2, так что (29) согласуется с более ранним определением (1.4.16), а именно $\delta=\delta_2-\delta_1$.

где q и χ — постоянные; тогда

$$\left.\begin{aligned} J_{xx} &= \langle a_1^2\rangle, & J_{xy} &= q\langle a_1^2\rangle \exp(i\chi), \\ J_{yx} &= q\langle a_1^2\rangle \exp(-i\chi), & J_{yy} &= q^2\langle a_1^2\rangle, \end{aligned}\right\} \tag{37}$$

и условие (30) выполняется. Матрица когерентности с элементами (37) совпадает с матрицей для строго монохроматического света с компонентами

$$E_x = \sqrt{\langle a_1^2\rangle} \exp\{i[\alpha - 2\pi\bar{\nu}t]\}, \quad E_y = q\sqrt{\langle a_1^2\rangle} \exp\{i[-\chi + \alpha - 2\pi\bar{\nu}t]\}, \tag{38}$$

где α — произвольная постоянная. Отсюда следует, что в эксперименте с поляризатором и компенсатором поведение квазимонохроматической волны, подчиняющейся условиям (36), в точности совпадает с поведением строго монохроматической и, следовательно, полностью поляризованной волны (38). (Предполагается, конечно, что разность фаз, вносимая компенсатором, мала по сравнению с длиной когерентности света, измеренной в единицах средней длины волны.) Поэтому можно сказать, что условие (30) характеризует *полностью поляризованную* световую волну.

10.8.2. Некоторые эквивалентные представления. Степень поляризации световой волны. При суперпозиции нескольких *независимых* световых волн, распространяющихся в одном направлении, матрица когерентности результирующей волны равна сумме матриц когерентности для отдельных волн. Чтобы доказать это, рассмотрим компоненты электрических векторов (в обычном комплексном представлении) отдельных волн $E_x^{(n)}$, $E_y^{(n)}$ $(n = 1, 2, \ldots, N)$. Компоненты электрического вектора результирующей волны равны

$$E_x = \sum_{n=1}^{N} E_x^{(n)}, \quad E_y = \sum_{n=1}^{N} E_y^{(n)}, \tag{39}$$

а значит, элементы матрицы когерентности для этой волны определяются выражениями

$$J_{kl} = \langle E_k E_l^*\rangle = \sum_{n=1}^{N}\sum_{m=1}^{N} \langle E_k^{(n)} E_l^{(m)*}\rangle = \sum_n \langle E_k^{(n)} E_l^{(n)*}\rangle + \sum_{n\neq m} \langle E_k^{(n)} E_l^{(m)*}\rangle. \tag{40}$$

Так как предполагается, что волны независимы, каждый член последней суммы равен нулю и, следовательно,

$$J_{kl} = \sum_n J_{kl}^{(n)}, \tag{41}$$

где $J_{kl}^{(n)} = \langle E_k^{(n)} E_l^{(n)*}\rangle$ — элементы матрицы когерентности n-й волны. Уравнение (41) показывает, что матрица когерентности для сложной волны равна сумме матриц когерентности для всех составляющих волн.

Вместе с тем любую волну можно рассматривать как сумму независимых волн, которые, очевидно, можно выбирать различными способами. Ниже мы кратко остановимся на одном специальном способе такого выбора.

Покажем, что любую квазимонохроматическую световую волну можно рассматривать как сумму полностью неполяризованной и полностью поляризованной волн, не зависящих друг от друга, и что такое представление единственно. Для этого необходимо лишь показать, что любую матрицу когерентности $\mathbf{J}$ можно единственным образом выразить в виде

$$\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)}, \tag{42}$$

где в соответствии с (27) и (30)

$$\mathbf{J}^{(1)} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}, \quad \mathbf{J}^{(2)} = \begin{bmatrix} B & D \\ D^* & C \end{bmatrix}, \tag{43}$$

причем $A \geqslant 0$, $B \geqslant 0$, $C \geqslant 0$ и

$$BC - DD^* = 0. \tag{44}$$

Если J_{xx}, J_{xy}, ...,— элементы матрицы когерентности, характеризующей

исходную волну, то на основании (42) и (43) имеем

$$\left.\begin{aligned} A+B&=J_{xx}, & D&=J_{xy},\\ D^*&=J_{yx}, & A+C&=J_{yy}. \end{aligned}\right\} \tag{45}$$

Подставляя (45) в (44), мы получим следующее уравнение для A:

$$(J_{xx}-A)(J_{yy}-A)-J_{xy}J_{yx}=0; \tag{46}$$

таким образом A является характеристическим корнем (собственным значением) матрицы когерентности J. Два корня уравнения (46) равны

$$A=\frac{1}{2}(J_{xx}+J_{yy})\pm\frac{1}{2}\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}, \tag{47}$$

где, как и раньше, $|\mathrm{J}|$ — определитель (8). Так как $J_{yx}=J^*_{xy}$, произведение $J_{xy}J_{yx}$ неотрицательно и из (8) следует, что

$$|\mathrm{J}|\leqslant J_{xx}J_{yy}\leqslant\frac{1}{4}(J_{xx}+J_{yy})^2;$$

значит оба корня (47) вещественны и неотрицательны. Рассмотрим вначале решение со знаком минус перед квадратным корнем. Имеем

$$A=\frac{1}{2}(J_{xx}+J_{yy})-\frac{1}{2}\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}, \tag{48}$$

$$\left.\begin{aligned} B&=\frac{1}{2}(J_{xx}-J_{yy})+\frac{1}{2}\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}, \quad D=J_{xy},\\ D^*&=J_{yx}, \quad C=\frac{1}{2}(J_{yy}-J_{xx})+\frac{1}{2}\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}. \end{aligned}\right\} \tag{49}$$

Поскольку

$$\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}=\sqrt{(J_{xx}-J_{yy})^2+4J_{xy}J_{yx}}\geqslant|J_{xx}-J_{yy}|,$$

B и C также неотрицательны, как и требуется. Второй корень (47) (со знаком плюс перед квадратным корнем) дает отрицательные значения B и C, и поэтому его следует отбросить. Таким образом, мы получили единственное разложение требуемого вида.

Полная интенсивность волны равна

$$I_{\text{полн}}=\mathrm{Sp}\,\mathrm{J}=J_{xx}+J_{yy}, \tag{50}$$

а полная интенсивность поляризованной части —

$$I_{\text{поляр}}=\mathrm{Sp}\,\mathrm{J}^{(2)}=B+C=\sqrt{(J_{xx}+J_{yy})^2-4|\mathrm{J}|}. \tag{51}$$

Отношение интенсивности поляризованной части к полной интенсивности называется *степенью поляризации* P волны; согласно (50) и (51) она определяется соотношением

$$P=\frac{I_{\text{поляр}}}{I_{\text{полн}}}=\sqrt{1-\frac{4|\mathrm{J}|}{(J_{xx}+J_{yy})^2}}. \tag{52}$$

Это выражение содержит лишь два инварианта вращения матрицы когерентности, и поэтому, как и следовало ожидать, степень поляризации не зависит от выбора осей Ox, Oy. Из (52) и неравенства, предшествующего (48), вытекает, что

$$0\leqslant P\leqslant 1. \tag{53}$$

Когда $P=1$, неполяризованная компонента отсутствует и, значит, волна *полностью поляризована*. При этом $|\mathrm{J}|=0$, так что $|\mu_{xy}|=1$ и, следовательно, E_x и E_y взаимно когерентны. Когда $P=0$, отсутствует поляризованная компонента. Волна тогда *полностью неполяризована*. В этом случае $(J_{xx}+J_{yy})^2=4|\mathrm{J}|$, т. е.

$$(J_{xx}-J_{yy})^2+4J_{xy}J_{yx}=0. \tag{54a}$$

Так как $J_{yx}=J_{xy}^*$, то находим, что равна нулю сумма квадратов двух величин, а это возможно лишь в том случае, когда каждая из них равна нулю, т. е. когда

$$J_{xx}=J_{yy} \quad \text{и} \quad J_{xy}=J_{yx}=0, \tag{54б}$$

что соответствует (26б). Тогда E_x и E_y взаимно некогерентны ($\mu_{xy}=0$). Во всех других случаях ($0<P<1$) мы говорим, что свет *частично поляризован.* Сравнение (52) и (20) показывает, что величина $\frac{(I(\theta, \varepsilon))_{\text{макс}}-(I(\theta, \varepsilon))_{\text{мин}}}{(I(\theta, \varepsilon))_{\text{макс}}+(I(\theta, \varepsilon))_{\text{мин}}}$ в точности равна степени поляризации P.

Выражение для степени поляризации принимает простую форму, когда E_x и E_y взаимно некогерентны (но свет не обязательно естественный). Так как в этом случае $J_{xy}=J_{yx}=0$, то $|\mathsf{J}|=J_{xx}J_{yy}$ и (52) переходит в *)

$$P=\left|\frac{J_{xx}-J_{yy}}{J_{xx}+J_{yy}}\right|. \tag{55}$$

Это выражение согласуется с формулой (1.5.42), используемой для определения поляризации естественного света при отражении.

Укажем несколько полезных представлений *естественного света.* Матрицу когерентности (27) для естественного света всегда можно записать в виде

$$\frac{1}{2}I\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\frac{1}{2}I\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}+\frac{1}{2}I\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}, \tag{56}$$

а это означает, согласно (33), что волна естественного света интенсивности I эквивалентна двум независимым линейно поляризованным волнам с интенсивностью каждой, равной $\frac{1}{2}I$, и электрическими векторами, колеблющимися в двух взаимно перпендикулярных направлениях, нормальных к направлению распространения.

Другое полезное представление естественного света имеет вид

$$\frac{1}{2}I\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\frac{1}{4}I\begin{bmatrix}1 & +i\\ -i & 1\end{bmatrix}+\frac{1}{4}I\begin{bmatrix}1 & -i\\ +i & 1\end{bmatrix}. \tag{57}$$

Согласно (35) оно означает, что волна естественного света интенсивности I эквивалентна двум независимым циркулярно поляризованным волнам с интенсивностью каждой, равной $\frac{1}{2}I$, причем одна из волн поляризована по правому кругу, другая — по левому.

Возвращаясь к общему случаю (частично поляризованный свет), следует отметить, что, в отличие от степени поляризации P, степень когерентности $|\mu_{xy}|$ зависит от выбора осей x и y. Однако легко показать, что $|\mu_{xy}|$ не может превышать P. Действительно, если в (52) мы запишем полное выражение детерминанта $|\mathsf{J}|$ и используем (6), то найдем, что

$$1-P^2=\frac{J_{xx}J_{yy}}{\left[\frac{1}{2}(J_{xx}+J_{yy})^2\right]}\,[1-|\mu_{xy}|^2]. \tag{58}$$

Так как среднее геометрическое любых двух положительных чисел не может превышать их среднего арифметического, то $1-P^2\leqslant 1-|\mu_{xy}|^2$, т. е.

$$P\geqslant|\mu_{xy}|. \tag{59}$$

*) Поскольку каждую эрмитову матрицу можно привести к диагональному виду с помощью унитарного преобразования, причем $|\mathsf{J}|$ и $\operatorname{Sp}\mathsf{J}$ инварианты при таком преобразовании, степень поляризации всегда можно представить в виде

$$P=\left|\frac{A_1-A_2}{A_1+A_2}\right|,$$

где A_1 и A_2 — два собственных значения (определяемых выражениями (47)). Однако, вообще говоря, унитарное преобразование не представляет действительного вращения осей вокруг направления распространения волны. Интересно отметить, что собственные значения A_1 и A_2 равны значениям $(I(\theta, \varepsilon))_{\text{макс}}$ и $(I(\theta, \varepsilon))_{\text{мин}}$, определяемым выражениями (19).

Знак равенства в (59) справедлив тогда и только тогда, когда $J_{xx}=J_{yy}$, т. е. когда интенсивности (средние по времени) x- и y-компонент электрического вектора одинаковы. Ниже мы покажем, что всегда существует пара направлений, для которых это справедливо.

При повороте осей x и y в плоскости, в которой они лежат, на угол θ против часовой стрелки, J_{xx} и J_{yy} переходят соответственно в $J_{x'x'}$ и $J_{y'y'}$, причем на основании (23)

$$\left.\begin{aligned} J_{x'x'} &= J_{xx}\cos^2\theta + J_{yy}\sin^2\theta + (J_{xy}+J_{yx})\cos\theta\sin\theta,\\ J_{y'y'} &= J_{xx}\sin^2\theta + J_{yy}\cos^2\theta - (J_{xy}+J_{yx})\cos\theta\sin\theta. \end{aligned}\right\} \qquad (60)$$

Из (60) следует, что $J_{x'x'}=J_{y'y'}$, если оси поворачиваются на угол $\theta=\Theta$, задаваемый соотношением

$$\operatorname{tg} 2\Theta = \frac{J_{yy}-J_{xx}}{J_{xy}+J_{yx}}. \qquad (61)$$

Так как $J_{yx}=J_{xy}^*$, а J_{xx} и J_{yy} вещественны, уравнение (61) всегда имеет вещественное решение для Θ. Таким образом, *всегда существует пара взаимно ортогональных направлений, для которых интенсивности равны. Для этой пары направлений степень когерентности $|\mu_{xy}|$ электрических колебаний принимает максимальное значение, равное P — степени поляризации волны* *).

10.8.3. Параметры Стокса квазимонохроматической плоской волны. Как мы видели, для характеристики квазимонохроматической плоской волны, вообще говоря, необходимы четыре вещественные величины, например J_{xx}, J_{yy}, вещественная и мнимая части J_{xy} (или J_{yx}). В своих исследованиях, относящихся к частично поляризованному свету, Стокс [27] ввел **) несколько отличное представление с четырьмя параметрами, тесно связанное с рассмотренным выше. Мы уже сталкивались с ним (в частном случае) в п. 1.4.2 при изучении монохроматического света. Параметрами Стокса общего вида являются следующие четыре величины:

$$\left.\begin{aligned} s_0 &= \langle a_1^2\rangle + \langle a_2^2\rangle, & s_1 &= \langle a_1^2\rangle - \langle a_2^2\rangle,\\ s_2 &= 2\langle a_1 a_2\cos\delta\rangle, & s_3 &= 2\langle a_1 a_2\sin\delta\rangle, \end{aligned}\right\} \qquad (62)$$

где, как и прежде, a_1 и a_2 — мгновенные амплитуды двух взаимно перпендикулярных компонент электрического вектора E_x и E_y, а $\delta=\varphi_1-\varphi_2$ — разность их фаз. Для монохроматического света a_1, a_2 и δ не зависят от времени, и выражения (62) переходят в параметры Стокса монохроматической волны, определенные в (1.4.43).

Из (62) и (4) следует, что параметры Стокса и элементы матрицы когерентности связаны соотношениями

$$\left.\begin{aligned} s_0 &= J_{xx}+J_{yy}, & s_1 &= J_{xx}-J_{yy},\\ s_2 &= J_{xy}+J_{yx}, & s_3 &= i(J_{yx}-J_{xy}); \end{aligned}\right\} \qquad (63а)$$

$$\left.\begin{aligned} J_{xx} &= \tfrac{1}{2}(s_0+s_1), & J_{yy} &= \tfrac{1}{2}(s_0-s_1),\\ J_{xy} &= \tfrac{1}{2}(s_2+is_3), & J_{yx} &= \tfrac{1}{2}(s_2-is_3). \end{aligned}\right\} \qquad (63б)$$

Как и элементы матрицы когерентности, параметры Стокса любой плоской квазимонохроматической волны можно определить с помощью простых экспериментов. Как и раньше, обозначим через $I(\theta, \varepsilon)$ интенсивность световых колебаний в направлении, образующем угол θ с осью Ox, когда их y-компонента

*) Геометрическая интерпретация этой специальной пары направлений рассмотрена Вольфом [82].

**) См. также [16, 28, 31, 34, 83, 84]. Параметры Стокса применяются и при квантовомеханическом рассмотрении поляризации элементарных частиц [85—88], см. также [30].

запаздывает на величину ε по отношению к x-компоненте. Тогда на основании соотношений (11) и (63а) имеем

$$\left.\begin{aligned} s_0 &= I(0^\circ, 0) + I(90^\circ, 0), & s_1 &= I(0^\circ, 0) - I(90^\circ, 0), \\ s_2 &= I(45^\circ, 0) - I(135^\circ, 0), & s_3 &= I\left(45^\circ, \frac{\pi}{2}\right) - I\left(135^\circ, \frac{\pi}{2}\right). \end{aligned}\right\} \quad (64)$$

Параметр s_0, очевидно, представляет полную интенсивность. Параметр s_1 равен разности интенсивностей линейно поляризованного света, прошедшего через поляризаторы, с азимутами $\theta = 0^\circ$ и $\theta = 90^\circ$. Так же интерпретируется и параметр s_2, но для азимутов $\theta = 45^\circ$ и $\theta = 135^\circ$. Наконец, параметр s_3 равен разности интенсивностей света, прошедшего через прибор, пропускающий колебания с правой круговой поляризацией, и света, прошедшего через прибор, пропускающий колебания с левой круговой поляризацией.

Если использовать соотношения (63б), то все приведенные выше результаты можно выразить не через матрицу когерентности, а с помощью параметров Стокса. В частности, условие (8), а именно $J_{xx}J_{yy} - J_{xy}J_{yx} \geqslant 0$, примет вид

$$s_0^2 \geqslant s_1^2 + s_2^2 + s_3^2. \quad (65)$$

Для монохроматического света имеем, согласно (30), $J_{xx}J_{yy} - J_{xy}J_{yx} = 0$, и тогда в соответствии с (1.4.44) в соотношении (65) мы получим знак равенства.

Рассмотрим теперь разложение данной волны на взаимно независимые поляризованную и неполяризованную части, используя представление через параметры Стокса. Из (41) и (63) следует, что параметры Стокса системы *независимых* волн равны сумме соответствующих параметров Стокса отдельных волн. Из (27) и (63а) вытекает, что для неполяризованной волны (естественный свет) справедливо соотношение $s_1 = s_2 = s_3 = 0$. Обозначим четыре параметра Стокса s_0, s_1, s_2, s_3 одним символом $\mathbf{s}$. Тогда, очевидно, для волны, характеризующейся параметром $\mathbf{s}$, требуемое разложение запишется в виде

$$\mathbf{s} = \mathbf{s}^{(1)} + \mathbf{s}^{(2)}, \quad (66)$$

где

$$\mathbf{s}^{(1)} = s_0 - \sqrt{s_1^2 + s_2^2 + s_3^2},\ 0,\ 0,\ 0, \quad (67а)$$

$$\mathbf{s}^{(2)} = \sqrt{s_1^2 + s_2^2 + s_3^2},\ s_1,\ s_2,\ s_3. \quad (67б)$$

Параметр $\mathbf{s}^{(1)}$ соответствует неполяризованной, а $\mathbf{s}^{(2)}$— поляризованной части волны. Следовательно, с помощью параметров Стокса степень поляризации исходной волны можно выразить в виде

$$P = \frac{I_{\text{поляр}}}{I_{\text{полн}}} = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0}. \quad (68)$$

То же выражение нетрудно получить, подставляя (63б) в (52). Легко также записать выражения, определяющие форму и ориентацию эллипса поляризации, связанного с поляризованной частью волны (см. (67б)). Если, как и в (1.4.28), соотношение $\operatorname{tg}\chi = \pm b/a$ $(-\pi/4 < \chi \leqslant \pi/4)$ определяет отношение малой и большой осей и направление, в котором описывается эллипс ($\chi > 0$ соответствует правой, а $\chi < 0$ — левой поляризации), то на основании (1.4.45в) и (67б) имеем

$$\sin 2\chi = \frac{s_3}{\sqrt{s_1^2 + s_2^2 + s_3^2}}. \quad (69)$$

Угол $\psi\ (0 \leqslant \psi < \pi)$ между большой осью и Ox определяется в соответствии с (1.4.46) и (67б) соотношением

$$\operatorname{tg} 2\psi = s_2/s_1. \quad (70)$$

Итак, мы видим, что параметры Стокса, как и матрица когерентности, служат полезным инструментом для систематического анализа состояния поляризации квазимонохроматической волны.

ЛИТЕРАТУРА

1. E. Verdet, Ann. Sci. École Normale Supérieure 2, 291 (1865); Leçons d'Optique Physique, L'Imprimerie Impériale, Paris, 1869, vol. 1, p. 106.
2. A. A. Michelson, Phil. Mag. 30, 1 (1890); 31, 256 (1891), Astrophys. J. 51, 257 (1920).
3. A. A. Michelson, Phil. Mag. 31, 338 (1891); 34, 280 (1892).
4. F. Zernike, Proc. Phys. Soc. 61, 158 (1948).
5. M. Laue, Ann. d. Physik 23, 1, 795 (1907).
6. M. Berek, Z. Phys. 36, 675, 824 (1926); 37, 387 (1926); 40, 420 (1926).
7. C. Lakeman, J. Th. Groosmuller, Physica 8, 193, 199, 305 (1928).
8. P. H. van Cittert, Physica 1, 201 (1934).
9. P. H. van Cittert, Physica 6, 1129 (1939).
10. L. Jánossy, Nuovo Cimento 6, 111 (1957); 12, 369 (1959).
11. F. Zernike, Physica 5, 785 (1938).
12. H. H. Hopkins, Proc. Roy. Soc. A208, 263 (1951); A217, 408 (1953).
13. D. Gabor, в сб. «Proc. Symp. Astr. Optics», ed. Z. Kopal, North Holland Publ. Co., Amsterdam, 1956, p. 17; в сб. «Proc. Third Symp. Inform. Theory», ed. C. Cherry, Butterworths Sci. Publ., London, 1956, p. 26.
14. H. Gamo, J. Appl. Phys. Japan 25, 431 (1956); в сб.: «Progr. in Optics», vol. 3, ed. E. Wolf, North Holland Publ. Co., Amsterdam, J. Wiley a. Sons, N. Y., 1964, p. 187.
15. E. Wolf, Proc. Roy. Soc. A230, 246 (1955); A225, 96 (1954).
16. E. Wolf, Nuovo Cimento 12, 884 1954).
17. A. Blanc-Lapierre, P. Dumontet, Rev. d'Opt. 34, 1 (1955).
18. J. J. Freeman, Principles of Noise, J. Wiley a. Sons, Inc., N. Y., 1958, p. 245—247.
19. E. Wolf, в сб. «Quantum Electronics», 3rd Congress, ed. N. Bloembergen a. P. Grivet, Columbia Univ. Press, N. Y., Dunod, Paris, 1964, p. 13.
20. L. Mandel, в сб.: «Quantum Electronics», 3rd Congress, ed. N. Bloembergen a. P. Grivet, Colubmia Univ. Press, N. Y., Dunod, Paris, 1964, p. 101.
21. E. Wolf, в сб.: «Proc. Symp. on Optical Masers», Brooklyn Polytech. Press, J. Wiley a. Sons, Inc. N. Y., 1963.
22. R. J. Glauber, в сб.: «Quantum Electronics», 3rd Congress, ed. N. Bloembergen a. P. Grivet, Columbia Univ. Press, N. Y., Dunod, Paris, 1964, p. 111; Phys. Rev. 130, 2529 (1963).
23. E. C. G. Sudarshan, Phys. Rev. Lett. 10, 277 (1963); J. R. Klauder, E. C. J. Sudarshan, Fundamentals of Quantum Optics, W. A. Benjamin, Inc., N. Y., 1968. (Дж. Клаудер, Э. Сударшан, Основы квантовой оптики, «Мир», 1971.)
24. L. Mandel, E. Wolf, Rev. Mod. Phys. 37, 231 (1965). (Э. Вольф, Л. Мандель, УФН 87, 491 (1965); 88, 347, 619 (1966).)
25. R. N. Bracewell, Radio Astronomy Techniques, в сб. «Encyclopedia of Physics», ed. S. Flügge, Springer, Berlin, 1959, vol. 54.
26. J. A. Ratcliffe, в сб. «Rep. Progr. Phys.», Phys. Soc. London, 1956, vol. 19, p. 188.
27. G. G. Stokes, Trans. Cambr. Phil. Soc. 9, 399 (1852); Mathematical and Physical Papers, Cambr. Univ. Press, 1901, vol. III, p. 233.
28. P. Soleillet, Ann. de Physique 12, 23 (1929).
29. N. Wiener, J. Math. Phys. 7, 109 (1928); J. Franklin Inst. 207, 525 (1929).
30. N. Wiener, Acta Math 55, § 9 (1930).
31. F. Perrin, J. Chem. Phys. 10, 415 (1942).
32. E. Wolf, в сб. «Proc. Symp. Astr. Optics», ed. Z. Kopal, North Holland Publ. Co., Amsterdam, 1956, p. 177.
33. E. Wolf, Nuovo Cimento 13, 1165 (1959).
34. S. Pancharatnam, Proc. Indian Acad. Sci. A44, 398 (1956); 57, 218, 231 (1963).
35. E. C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Clarendon Press, Oxford, 2nd ed., 1948. (E. Титчмарш, Введение в теорию интегралов Фурье, Гостехиздат, 1948.)
36. D. Gabor, J. IEE 93, Pt. III, 429 (1946).
37. В. И. Бунимович, ЖТФ 19, 1231 (1949).
38. J. Ville, Câbles et Transmission 2, 61 (1948); 4, 9 (1950).
39. J. R. Oswald, Trans. IRE, CT-3, 244 (1956).
40. H. A. Kramers, Atti Congr. Intern. Fisici, Como (Sept. 1927), (V), N. Zanichelli, Bologna, 1928.
41. J. S. Toll, Phys. Rev. 104, 1760 (1956), J. Hilgevoord, Dispersion Relations and Causal Description, North Holland Publ Co., Amsterdam, 1960.
42. L. Mandel, J. Opt. Soc. Amer. 57, 613 (1967).
43. N. Wiener, Acta Math. 55, 117 (1930).
44. W. B. Davenport, W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, N. Y., 1958. (В. Б. Давенпорт, В. Л. Рут, Введение в теорию случайных сигналов и шумов, ИЛ, 1960.)
45. D. Middleton, Trans, IRE CT-3, 299 (1956).

46. S. Goldman, Information Theory, Prentice-Hall, Inc., N. Y., 1953. (С. Голдман, Теория информации, ИЛ, 1957.)
47. А. М. Яглом, Введение в теорию стационарных случайных функций, УМН 7, 3 (1952).
48. J. L. Doob, Stochastic Processes, J. Wiley a. Sons, Inc., New York, 1953
49. M. S. Bartlett, An Introduction to Stochastic Processes, Cambr. Univ. Press, 1955.
50. H. Margenau, G. M. Murphy, The Mathematics of Physics and Chemistry, D. van Nostrand Co., Amsterdam, 1947.
51. L. Mandel, E. Wolf, J. Opt. Soc. Amer. 51, 815 (1961).
52. A. Khintchine, Math. Ann. 109, 604 (1934). (А. Я. Хинчин, УМН вып. 5, 42 (1938).)
53. P. Roman, E. Wolf, Nuovo Cimento 17, 474—476 (1960).
54. L. Mandel, в сб. «Progr. in Optics», ed. E. Wolf, North Holland Publ. Co., Amsterdam, 1963, vol. 2, p. 181.
55. R. H. Brown, R. Q. a. Twiss, Proc. Roy. Soc. A242, 300 (1957); A243, 291 (1957).
56. E. M. Purcell, Nature 178, 1449 (1956).
57. F. D. Kahn, Optica Acta 5, 93 (1958).
58. L. Mandel, Proc. Phys. Soc. 72, 1037 (1958).
59. E. Wolf, Phil. Mag. 2, 351 (1957).
60. J. A. Ratcliffe, в сб. «Rep. Progr. Phys.», Phys. Soc., 1956, vol. 19, p. 233.
61. C. A. Taylor, R. M. Hinde, H. Lipson, Acta Cryst. 4, 261 (1951).
62. A. W. Hanson, H. Lipson, C. A. Taylor, Proc. Roy. Soc. A218, 371 (1953).
63. W. Hughes, C. A. Taylor, J. Sci. Instr. 30, 105 (1953).
64. A. T. Forrester, Amer. J. Phys. 24, 194 (1956).
65. J. Thompson, E. Wolf, J. Opt. Soc. Amer. 47, 895 (1957).
66. F. Zernike, Physica 5, 791 (1938).
67. F. Zernike, Physica 5, 794 (1938).
68. H. H. Hopkins, P. M. Barham, Proc. Phys. Soc. 63, 72 (1950).
69. A. Köhler, Z. f. wiss. Mikrosk. 10, 433 (1893); 16, 1 (1899).
70. P. Dumontet, Publ. Sci. Univ. d'Alger B1, 33 (1955).
71. E. Wolf, Proc. Phys. Soc. 71, 257 (1958).
72. M. J. Beran, G. B. Parrent, Theory of Partial Coherence, Englewood Cliffs, N. Y., 1964, § 3.3.
73. L. Mandel, Proc. Phys. Soc. 74, 233 (1959); L. Mandel, E. Wolf, Proc. Phys. Soc. 80, 894 (1962).
74. R. Silverman, Trans. IRE CT-5, 84 (1958).
75. А. Г. Майер, Е. А. Леонтович, ДАН СССР 4, 353 (1934).
76. I. Kay, R. Silverman, Information and Control 1, 64, 396 (1957).
77. А. А. Харкевич, Спектры и анализ, Физматгиз, 1962, § 12.
78. H. Weyl, The Theory of Groups and Quantum Mechanics, Methuen, London, 1931.
79. G. B. Parrent, P. Roman, Nuovo Cimento 15, 370 (1960).
80. P. Roman, E. Wolf, Nuovo Cimento 17, 462, 477 (1960).
81. P. Roman, Nuovo Cimento 20, 759 (1961); 22, 1005 (1961).
82. E. Wolf, Nuovo Cimento 13, 1180—1181 (1959).
83. S. Chandrasekhar, Radiative Transfer, Clarendon Press, Oxford, 1950, § 15.
84. M. J. Walker, Amer. J. Phys. 22, 170 (1954).
85. U. Fano, J. Opt. Soc. Amer. 39, 859 (1949); 41, 58 (1951); Phys. Rev. 93, 121 (1954).
86. D. L. Falkoff, J. E. Mac Donald, J. Opt. Soc. Amer. 41, 861 (1951).
87. W. H. MacMaster, Amer. J. Phys. 22, 351 (1954).
88. J. M. Jauch, F. Rohrlich, The Theory of Photons and Electrons, Addison-Wesley Publ. Co. Inc., Cambridge, 1955, § 2.8.
89*. T. Yamamoto, Opt. Acta 12, 229 (1965).
90*. Y. O'hmen, Appl. Opt. 4, 1660 (1965).
91*. E. L. O'Neill, Introduction to statistical Optics, Addison-Wesley Publ. Co., London, 1963. (Э. О'Нейл, Введение в статистическую оптику, «Мир», 1966.)
92*. С. М. Рытов, Введение в статистическую радиофизику, «Наука», 1966.
93*. Р. Глаубер, в сб. «Квантовая оптика и квантовая радиофизика», «Мир», 1971.

ГЛАВА 11

СТРОГАЯ ТЕОРИЯ ДИФРАКЦИИ

§ 11.1. Введение

Если исходить из уравнений Масквелла и обычных граничных условий, то проблема дифракции электромагнитного излучения на некотором теле сводится к строго определенной краевой математической задаче. В настоящей главе некоторые аспекты теории дифракции монохроматических волн обсуждаются именно с этой точки зрения. В частности, здесь подробно и строго решается классическая задача дифракции на идеально проводящей полуплоскости.

В ранних теориях, принадлежащих Юнгу, Френелю и Кирхгофу, дифракционное препятствие предполагалось абсолютно «черным»; иными словами, считалось, что все излучение, падающее на препятствие, полностью поглощается, не отражаясь. Это предположение послужило источником внутренних противоречий, так как такое понятие абсолютной «черноты» нельзя было точно определить и оно явно было несовместимым с электромагнитной теорией.

Случаи, когда тело, на котором происходит дифракция, имеет конечную диэлектрическую проницаемость и конечную проводимость, исследовались теоретически; одно из первых исчерпывающих исследований такого рода для дифракции на сфере, выполненное в 1908 г. Ми, рассматривается в гл. 13, посвященной оптике металлов. Вообще говоря, предположение о конечной проводимости приводит к очень большому усложнению математического аппарата, и поэтому часто желательно принять концепцию идеально проводящего (и, следовательно, идеального отражающего) тела. Это, конечно, идеализация, но совместимая с электромагнитной теорией; кроме того, поскольку проводимость некоторых металлов (например, меди) очень велика, подобное представление может служить хорошей аппроксимацией, если частота не слишком велика. Однако следует подчеркнуть, что такая аппроксимация на оптических частотах никогда не является полностью адекватной. Упрощающее предположение о бесконечной проводимости дифракционного препятствия принято в большинстве работ, основанных на строгих математических выводах; наши последующие рассуждения также ограничиваются этим случаем.

Первое строгое решение такой дифракционной задачи было дано в 1896 г. Зоммерфельдом [1] при рассмотрении двумерного случая падения плоской волны на бесконечно тонкую идеально проводящую полуплоскость. Широкая известность его результата основана частично на том искусстве, с которым была решена задача, а частично на том, что найденное им решение можно было выразить точно и просто в виде интегралов Френеля, столь характерных для прежних приближенных теорий.

Многие математики последовали по пути Зоммерфельда. Ранние решения дифракционных задач, относящиеся к точечному и линейному источникам, а также интересные обобщения при рассмотрении дифракции на клине, а не на полуплоскости, связаны с именами Карслоу [2], Макдональда [3] и Бромвича [4]. Были решены также и другие задачи. В последнее время предложены новые методы, применение которых было стимулировано успехами ультракоротковолновой радиотехники. Прежде чем перейти к основному содержанию настоящей главы, упомянем очень коротко о некоторых из этих исследований.

Если существует такая ортогональная система координат u_1, u_2, u_3, в которой поверхность дифракционного тела совпадает с одной из поверхностей

u_i = const, то для решения дифференциальных уравнений в частных производных можно использовать классический способ разделения переменных. Таким методом фактически и воспользовался Ми для решения упоминавшейся выше задачи о сфере, обладающей конечной проводимостью. В этом случае решение краевой задачи имеет вид бесконечного ряда и его ценность зависит от легкости вычисления необходимых функций, а также от скорости, с которой ряд сходится. Этот метод применялся в различных случаях (помимо задачи со сферой); особенно надо отметить его использование в случае дифракции на круглом диске или отверстии [5]. Следует, однако, заметить, что лишь некоторые из этих работ относятся к чисто скалярным задачам типа задач, встречающихся в теории звуковых волн малой амплитуды; дальше будет показано, что двумерные задачи в электромагнитной теории принадлежат в основном к этому типу, но в других случаях векторная природа электромагнитного поля приводит к дополнительным осложнениям.

Другой метод с использованием интегральных уравнений, по-видимому, впервые рассматривался Рэлеем [6]. Некоторые задачи (простейшие из них относятся к полуплоскости) приводят к таким интегральным уравнениям, которые можно точно решить методом, развитым Винером и Хопфом *). Его использование Копсоном [8], Швингером и другими, дало ряд новых решений в замкнутой форме [9—11] (более подробная библиография указана в [12, 13]). В этой связи следует упомянуть также о мощном, хотя и несколько сложном вариационном методе, которым можно воспользоваться при расчете энергии, дифрагирующей через отверстие [14].

Из-за ограниченности места мы рассмотрим в настоящей главе только один метод **). Сначала мы изложим некоторые соображения общего характера, имеющие значение в теории дифракции электромагнитных волн на идеально проводящих структурах. Далее введем представление произвольного поля в виде интеграла по спектру плоских волн и покажем, что это ведет к формулировке некоторых дифракционных задач через «дуальные» интегральные уравнения ***). При этом задача Зоммерфельда с полуплоскостью легко решается; это решение приводится здесь и достаточно подробно исследуется вместе с некоторым числом побочных вопросов. В настоящей главе рассматривается также несколько смежных вопросов.

§ 11.2. Граничные условия и поверхностные токи

Хорошо известно, что электромагнитное поле только незначительно проникает в глубь хорошего проводника. Понятие бесконечной проводимости, при которой полностью отсутствует проникновение электромагнитного поля в глубь проводника, приводит, как покажут приведенные ниже аргументы, к понятию электрического тока, существующего исключительно на поверхности проводника.

Согласно уравнениям Максвелла (см. § 1.1.3) тангенциальная компонента **E** непрерывна при пересечении бесконечно тонкого листка электрического тока, тогда как аналогичная компонента **H** испытывает разрыв. Более точно, разрывна тангенциальная компонента **H**, нормальная к вектору поверхностной плотности****) **J** тока, и величина разрыва составляет $4\pi\mathbf{J}/c$. Относительное расположение этих векторов показано на рис. 11.1. Кроме того, в соответствии

*) Метод Винера и Хопфа разбирается в [7].

**) Изложение других методов дано в статье Вольфсона в [15] и гл. 4 и 5 книги [16] и в статье [17]. В обзоре [13] приведена наиболее полная сводка методов и формул.

***) «Дуальные» интегральные уравнения (они определяются ниже на стр. 520) рассматриваются в книге Титчмарша [7].

****) Поверхностный ток в п. 1.1.3 обозначен через **j**.

с поведением тангенциальных компонент **E** и **H** нормальная компонента **H** непрерывна при пересечении листка тока, тогда как нормальная компонента **E** разрывна и величина этого разрыва равна произведению поверхностной плотности зарядов на 4π. Таким образом очевидно, что поле в вакууме вне идеально проводящего тела таково, что на поверхности проводника имеем:

а) тангенциальная компонента **E** равна нулю;

б) тангенциальная компонента **H** перпендикулярна к вектору поверхностной плотности **J** электрического тока и равна $4\pi\mathbf{J}/c$;

в) нормальная компонента **H** равна нулю;

г) внешняя нормальная компонента **E** равна произведению 4π на поверхностную плотность заряда.

Действие излучения, падающего на идеально проводящее тело, удобно интерпретировать, выражая его через индуцированные поверхностные токи. Если $\mathbf{E}^{(i)}$ — электрический вектор падающего поля, а $\mathbf{E}^{(s)}$ — электрический вектор «рассеянного» поля, обусловленного индуцированным током, то полный электрический вектор повсюду равен $\mathbf{E}^{i}+\mathbf{E}^{(s)}$. Следовательно, теперь дифракционную задачу можно сформулировать следующим образом: *дано* $\mathbf{E}^{(i)}$, *требуется найти поле* $\mathbf{E}^{(s)}$, *создаваемое током, распределенным по поверхности проводника, которое имеет такую величину, что его тангенциальная составляющая на поверхности равна тангенциальной составляющей* $\mathbf{E}^{(i)}$, *взятой со знаком минус.* Стоит подчеркнуть, что приведенное выше граничное условие а) является основным и вместе с тем достаточным условием единственности решения вопроса лишь в том виде, как он сформулирован *). Что же касается других условий; то б) связывает поле с индуцированным током, а условия в) и г) не представляют особого интереса.

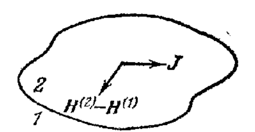


Рис. 11.1. Взаимное расположение векторов $\mathbf{H}^{(2)}-\mathbf{H}^{(1)}$ и **J**.

$\mathbf{H}^{(1)}$ и $\mathbf{H}^{(2)}$ — магнитные поля на первой и второй сторонах поверхности соответственно, **J** — вектор поверхностной плотности тока.

Следует отметить, что из равенства $\mathbf{E}^{(s)}=-\mathbf{E}^{(i)}$ во внутренних точках проводника вытекает существование на любой замкнутой поверхности S единственной плотности тока, которая обусловливает во всех точках внутри S поле, создаваемое источниками, находящимися вне S. Аналогично из рассмотрения случая, когда граница идеального проводника, на которую падает излучение, является бесконечной плоскостью, следует, что в любой плоскости существует единственная плотность тока, которая обусловливает на одной стороне этой плоскости поле, создаваемое источниками, расположенными на другой ее стороне.

В задачах о дифракции на идеально проводящих экранах существенно предположение, что они бесконечно тонки; если такое предположение не делается, то математические трудности значительно возрастают. Конечно, непрозрачность экрана по-прежнему подразумевается, причем это означает, что экран — идеальный проводник, толщина которого стремится в пределе к нулю. Из всего сказанного выше можно сделать вывод, что действие такого экрана мы вправе уподобить действию листка электрического тока, с той разницей, что теперь этот листок уже не является замкнутой поверхностью. Особый интерес представляет относительно простой случай плоского листка. В этом случае сразу же можно вывести некоторые важные соотношения, которым удовлетворяют излучаемые этим листком поля $\mathbf{E}^{(s)}$ и $\mathbf{H}^{(s)}$. Полученные соотношения приведены ниже.

---

*) Доказательство единственности представляет некоторые трудности и откладывается до § 11.9.

Допустим, что листок тока занимает часть плоскости $y=0$. Тогда из соображений симметрии ясно, что

$$\left.\begin{aligned} &E_x^{(s)}(x,\ y,\ z)=E_x^{(s)}(x,\ -y,\ z), && H_x^{(s)}(x,\ y,\ z)=-H_x^{(s)}(x,\ -y,\ z),\\ &E_y^{(s)}(x,\ y,\ z)=-E_y^{(s)}(x,\ -y,\ z), && H_y^{(s)}(x,\ y,\ z)=H_y^{(s)}(x,\ -y,\ z),\\ &E_z^{(s)}(x,\ y,\ z)=E_z^{(s)}(x,\ -y,\ z), && H_z^{(s)}(x,\ y,\ z)=-H_z^{(s)}(x,\ -y,\ z). \end{aligned}\right\} \qquad (1)$$

Кроме того, если компоненты плотности тока в листке обозначить через J_x и J_z, то, очевидно, при $y=0$ имеем

$$H_x^{(s)}=\mp\frac{2\pi}{c}J_z, \qquad H_z^{(s)}=\pm\frac{2\pi}{c}J_x. \qquad (2)$$

Верхний или нижний знак берется в зависимости от того, обращается ли y в нуль со стороны положительных или отрицательных значений. В следующем параграфе рассматриваются приложения этих простых соотношений к интересной задаче дифракции на плоском экране и выводится полезная формула, которая, в частности, служит доказательством существования точной электромагнитной аналогии принципа Бабине.

§ 11.3. Дифракция на плоском экране; электромагнитная форма принципа Бабине

Предположим, что электромагнитное поле $\mathbf{E}^{(i)}$, $\mathbf{H}^{(i)}$ падает на систему бесконечно тонких идеально проводящих пластинок, лежащих в плоскости $y=0$. Пусть M обозначает часть этой плоскости, занятую металлом, а A — оставшиеся «отверстия». Таким образом, M и A вместе составляют полную плоскость. Как каждая часть (M или A), так и обе они могут иметь бесконечные размеры.

Как указывалось выше, ищется дифрагировавшее поле, которое удовлетворяет определенным граничным условиям на M. Однако из (11.2.1) следует, что если известно в явном виде условие непрерывности при переходе через A, то необходимо рассмотреть дифрагировавшее поле только в одном из полупространств $y\geqslant 0$ или $y\leqslant 0$. Следовательно, нашу задачу можно сформулировать так: в полупространстве $y\geqslant 0$ (или $y\leqslant 0$) требуется найти электромагнитное поле $\mathbf{E}^{(s)}$, $\mathbf{H}^{(s)}$, создаваемое токами в плоскости $y=0$, для которого

(а) $\quad E_x^{(i)}+E_x^{(s)}=E_z^{(i)}+E_z^{(s)}=0$ на M;

(б) $\quad H_x^{(s)}=H_z^{(s)}=0$ на A.

Здесь (а) — основное граничное условие для идеального проводника, тогда как (б), следующее из (11.2.2),— удобный способ выразить отсутствие в A индуцированных токов. Если (б) удовлетворяется для дифрагировавшего поля в области $y\geqslant 0$, а для его определения в области $y\leqslant 0$ используется (11.2.1), то выполняется условие непрерывности при переходе через A.

Теперь легко вывести принцип Бабине для электромагнитных волн и идеально проводящих экранов в точном виде [18, 19]. Как и при классическом изложении принципа (см. п. 8.3.2), здесь устанавливается соотношение между соответствующими полями в присутствии основного экрана и «дополнительного» экрана, полученного путем замены проводящих тонких пластинок отверстиями; различие заключается в том, что падающее на дополнительный экран поле отлично от того, которое падает на основной экран, и получается из него преобразованием $\mathbf{E}\rightarrow\mathbf{H}$.

Пусть при $y>0$ поле (индекс 1), падающее на основной экран, равно $\mathbf{E}_1^{(i)}=\mathbf{F}^{(i)}$. Тогда из (а) и (б) находим:

(а′) $E_{1x}^{(s)}=-F_x^{(i)}, \quad E_{1z}^{(s)}=-F_z^{(i)}$ на M,

(б′) $H_{1x}^{(s)}=H_{1z}^{(s)}=0$ на A.

Пусть поле (индекс 2), падающее на дополнительный экран, равно $\mathbf{H}_2^{(i)}=-\mathbf{F}^{(i)}$. Теперь выразим граничные условия через полное поле:

(а″) $E_{2x}=E_{2z}=0$ на A,

(б″) $H_{2x}=F_x^{(i)}, \quad H_{2z}=F_z^{(i)}$ на M.

Так как уравнения Максвелла в вакууме инвариантны относительно преобразования $\mathbf{E}\to\mathbf{H}$, $\mathbf{H}\to-\mathbf{E}$ и так как имеется единственная поверхностная плотность тока в плоскости $y=0$, которая вызывала бы падающее поле во всех точках $y\leqslant 0$, то из сравнения (а′) и (б′) соответственно с (б″) и (а″) ясно, что в полупространстве *позади* экрана

$$\mathbf{H}_2=-\mathbf{E}_1^{(s)}. \tag{1}$$

Используя полное поле $\mathbf{E}_1$, получим из (1)

$$\mathbf{E}_1+\mathbf{H}_2=\mathbf{F}^{(i)}. \tag{2}$$

Это и есть электромагнитная форма принципа Бабине.

§ 11.4. Двумерная дифракция на плоском экране

11.4.1. Скалярная природа двумерных электромагнитных полей. Двумерной задачей называется такая задача, которая совершенно не зависит от какой-нибудь одной декартовой координаты, например от z. Как отмечалось выше, такие задачи в электромагнитной теории существенно скалярны, так как в них входит одна переменная. Сейчас это будет показано.

Отбросив временной множитель $\exp(-i\omega t)$ и написав $k=\omega/c$, можно привести уравнения Максвелла в вакууме к виду

$$\operatorname{rot}\mathbf{H}=-ik\mathbf{E}, \qquad \operatorname{rot}\mathbf{E}=ik\mathbf{H}.$$

Приравнивая нулю все частные производные по z, разделим эти уравнения на две независимые группы, а именно

$$\frac{\partial E_z}{\partial y}=ikH_x, \qquad \frac{\partial E_z}{\partial x}=-ikH_y, \qquad \frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=-ikE_z, \tag{1}$$

и

$$\frac{\partial H_z}{\partial y}=-ikE_x, \qquad \frac{\partial H_z}{\partial x}=ikE_y, \qquad \frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}=ikH_z. \tag{2}$$

В первую группу входят только H_x, H_y и E_z, во вторую — только E_x, E_y и H_z. Поэтому можно добиться упрощения, разделяя произвольное решение на линейную комбинацию двух решений, у которых каждый член одной из указанных выше групп равен нулю. Охарактеризуем эти два типа полей следующим образом:

E-поляризация

$$E_x=E_y=H_z=0, \qquad H_x=\frac{1}{ik}\frac{\partial E_z}{\partial y}, \qquad H_y=-\frac{1}{ik}\frac{\partial E_z}{\partial x},$$

и, как очевидно, подставляя H_x и H_y в третье уравнение (1), получим

$$\frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}+k^2E_z=0.$$

Здесь полное поле выражается через компоненту E_z, которая, конечно, удовлетворяет двумерному виду обычного волнового уравнения.

H-поляризация

$$H_x = H_y = E_z = 0$$

и

$$E_x = -\frac{1}{ik}\frac{\partial H_z}{\partial y}, \quad E_y = \frac{1}{ik}\frac{\partial H_z}{\partial x}.$$

Здесь полное поле выражается через H_z.

11.4.2. Угловой спектр плоских волн. Как было показано выше, в случае двумерных задач, которыми мы ограничиваемся, декартовые компоненты векторов **E** и **H** удовлетворяют уравнению

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + k^2 V = 0. \tag{3}$$

Это уравнение должно решаться при соответствующих граничных условиях. Основное элементарное решение уравнения (3) имеет вид

$$\exp[ikr\cos(\theta - \alpha)] = \exp[ik(x\cos\alpha + y\sin\alpha)], \tag{4}$$

где r и θ $(0 \leqslant \theta \leqslant 2\pi)$ — полярные координаты, связанные с x и y уравнениями $x = r\cos\theta$ и $y = r\sin\theta$, а α — угол между направлением распространения волны и осью x (рис. 11.2, *а*). При вещественном α выражение (4) представляет плоскую *однородную* волну, т. е. такую волну, у которой плоскости равных амплитуд и равных фаз совпадают. Если же α комплексно, выражение (4) представляет *неоднородную* плоскую волну, т. е. такую волну, у которой плоскости равных амплитуд и равных фаз не совпадают. Действительно, обозначив $\alpha = \alpha_1 + i\alpha_2$, где α_1 и α_2 вещественны, получим вместо (4)

Рис. 11.2. Плоская волна, описываемая соотношением (4).

а — однородная плоская волна при вещественном α, пунктирные линии обозначают плоскости равных амплитуд и равных фаз; *б* — неоднородная плоская волна при комплексном α, сплошные линии обозначают плоскости равных амплитуд, пунктирные линии — плоскости равных фаз

$$\exp[ikr\cos(\theta - \alpha)] = \exp[ikr\,\mathrm{ch}\,\alpha_2\cos(\theta - \alpha_1)]\exp[-kr\,\mathrm{sh}\,\alpha_2\sin(\theta - \alpha_1)]; \tag{5}$$

отсюда следует, что плоскости равных амплитуд и плоскости равных фаз взаимно перпендикулярны (рис. 11.2, *б*). Направление распространения фазы составляет с осью x угол α_1, причем фазовая скорость уменьшается в $\mathrm{sech}\,\alpha_2$ раз. В перпендикулярном направлении имеет место экспоненциальное ослабление, определяемое коэффициентом $k\,\mathrm{sh}\,\alpha_2$.

Теперь покажем [20], что при соответствующем выборе пути интегрирования и подходящем выборе функции $f(\alpha)$ любое решение (3) можно представить в виде *углового спектра плоских волн*

$$\int f(\alpha)\exp[ikr\cos(\theta - \alpha)]\,d\alpha.$$

Такое представление тесно связано с представлением произвольной функции с помощью интеграла Фурье и подобно последнему широко используется в приложениях. Без значительной потери в общности можно задать некоторый фиксированный путь интегрирования и тем самым свести любую задачу к отысканию соответствующей функции $f(\alpha)$. Представим сначала таким образом электромагнитное поле, создаваемое плоским листком тока, а затем покажем, что

это приведет нас к формулировке задачи дифракции на плоском экране в виде дуальных интегральных уравнений.

Рассмотрим двумерный листок тока в плоскости $y=0$. Как уже отмечалось выше, удобнее иметь дело отдельно с E-поляризацией и H-поляризацией. Сначала рассмотрим первый случай, когда плотность тока имеет только z-компоненту, допустим J_z. Попытаемся выяснить, при каком особом распределении появляется E-поляризованная плоская волна

$$\mathbf{E}=(0,0,1)\exp[ikr\cos(\theta-\alpha)],$$
$$\mathbf{H}=(\sin\alpha,\ -\cos\alpha,0)\exp[ikr\cos(\theta-\alpha)] \qquad (6)$$

в полупространстве $y>0$. Из первого соотношения (11.2.2) сразу же видно, что в точке $(\xi, 0)$

$$J_z(\xi)=-\frac{c}{2\pi}\exp(ik\xi\cos\alpha)\sin\alpha. \qquad (7)$$

Это конечно, можно проверить обычным методом потенциалов Герца, находя поле, создаваемое данным распределением тока, но тогда придется вычислять достаточно сложный интеграл.

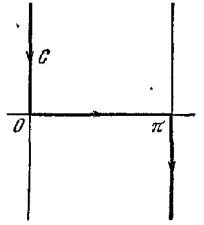


Рис. 11.3. Путь интегрирования C в комплексной α-плоскости.

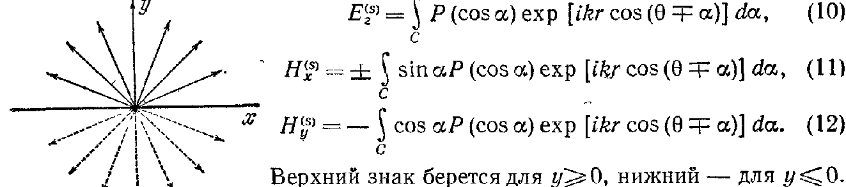
Итак, вообще говоря, можно создать любое распределение тока путем соответствующей суперпозиции выражений (7) для разных значений α, а возникающее поле можно получить соответствующей суперпозицией плоских волн (6). Более точно, допустим, что плотность тока можно записать в виде интеграла Фурье

$$J_z(\xi)=-\frac{c}{2\pi}\int_{-\infty}^{\infty}P(\mu)\exp(ik\xi\mu)\,d\mu. \qquad (8)$$

Замена переменной $\mu=\cos\alpha$ дает

$$J_z(\xi)=-\frac{c}{2\pi}\int_C \sin\alpha P(\cos\alpha)\exp(ik\xi\cos\alpha)\,d\alpha, \qquad (9)$$

где C — путь в комплексной α-плоскости, вдоль которого $\cos\alpha$ изменяется (принимая только вещественные значения) от ∞ до $-\infty$ (рис. 11.3). Результирующие не равные нулю компоненты поля равны, следовательно,



$$E_z^{(s)}=\int_C P(\cos\alpha)\exp[ikr\cos(\theta\mp\alpha)]\,d\alpha, \qquad (10)$$

$$H_x^{(s)}=\pm\int_C \sin\alpha P(\cos\alpha)\exp[ikr\cos(\theta\mp\alpha)]\,d\alpha, \qquad (11)$$

$$H_y^{(s)}=-\int_C \cos\alpha P(\cos\alpha)\exp[ikr\cos(\theta\mp\alpha)]\,d\alpha. \qquad (12)$$

Рис. 11.4. Направления распространения однородных волн, излучаемых в полупространство $y>0$ (сплошные линии) и в полупространство $y<0$ (пунктирные линии).

Верхний знак берется для $y\geqslant 0$, нижний — для $y\leqslant 0$.

Уравнения (10), (11) и (12) представляют поле в виде спектра плоских волн, определенного с помощью функции $P(\cos\alpha)$. Отдельные плоские волны, соответствующие части пути C вдоль вещественной оси, однородны и излучаются в области $y>0$ и $y<0$; в каждой области направления их распространения лежат в пределах угла π (рис. 11.4). Плоские волны, соответствующие двум частям C, для которых $\alpha=i\beta$ и $\alpha=\pi-i\beta$ (от $\beta=0$ до $\beta=\infty$), неоднородны; для них распространение фазы происходит вдоль положительного или отрицательного направления оси x, и имеет место экспоненциальное ослабление волны в направлении, нормальном к плоскости $y=0$. Легко показать,

исследуя поведение вектора Пойнтинга, что в среднем ни одна из этих *исчезающих* волн не переносит энергии от плоскости $y=0$. Наличие таких волн необходимо для учета структуры в распределении тока, которая мельче длины волны.

В случае H-поляризации поле, созданное током с плотностью тока J_x на $y=0$, можно аналогично записать в виде

$$H_z^{(s)} = \pm \int_C P(\cos\alpha) \exp[ikr\cos(\theta \mp \alpha)]\, d\alpha, \tag{13}$$

$$E_x^{(s)} = -\int_C \sin\alpha P(\cos\alpha) \exp[ikr\cos(\theta \mp \alpha)]\, d\alpha, \tag{14}$$

$$E_y^{(s)} = \pm \int_C \cos\alpha P(\cos\alpha) \exp[ikr\cos(\theta \mp \alpha)]\, d\alpha. \tag{15}$$

Верхний знак берется для $y \geqslant 0$, нижний — для $y \leqslant 0$, причем

$$J_x(\xi) = \frac{c}{2\pi} \int_{-\infty}^{\infty} \frac{P(\mu)}{\sqrt{1-\mu^2}} \exp(ik\xi\mu)\, d\mu. \tag{16}$$

11.4.3. Формулировка задачи через дуальные интегральные уравнения. Теперь можно сформулировать двумерную задачу дифракции на плоском экране через дуальные интегральные уравнения.

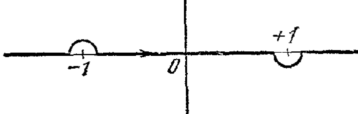


Рис. 11.5. Путь интегрирования от $-\infty$ до $+\infty$ в комплексной μ-плоскости.

Допустим, что электромагнитное поле $\mathbf{E}^{(i)}$, $\mathbf{H}^{(i)}$ падает на систему бесконечно тонких идеально проводящих полосок, лежащих в плоскости $y=0$; обозначим через M область изменения x, занятую металлом, а через A — область, свободную от металла. Если рассеянное поле $\mathbf{E}^{(s)}$, $\mathbf{H}^{(s)}$ представить в виде углового спектра плоских волн в форме (10), (11) и (12) или (13), (14) и (15) в соответствии с типом поляризации, то условия (а) и (б), приведенные в § 11.3, дают следующие интегральные уравнения:

E-поляризация

$$\int_{-\infty}^{\infty} \frac{P(\mu)}{\sqrt{1-\mu^2}} \exp(ikx\mu)\, d\mu = -E_z^{(i)} \quad \text{на} \quad M, \tag{17}$$

$$\int_{-\infty}^{\infty} P(\mu) \exp(ikx\mu)\, d\mu = 0 \quad \text{на} \quad A. \tag{18}$$

H-поляризация

$$\int_{-\infty}^{\infty} P(\mu) \exp(ikx\mu)\, d\mu = E_x^{(i)} \quad \text{на} \quad M, \tag{19}$$

$$\int_{-\infty}^{\infty} \frac{P(\mu)}{\sqrt{1-\mu}} \exp(ikx\mu)\, d\mu = 0 \quad \text{на} \quad A. \tag{20}$$

Анализ отображения части комплексной плоскости α от $\mathrm{Re}\,\alpha = 0$ до $\mathrm{Re}\,\alpha = \pi$ на всю комплексную плоскость μ ($\mu = \cos\alpha$) показывает, что путь интегрирования вдоль вещественной оси огибает возможные точки ветвления при $\mu = \pm 1$ (рис. 11.5). Интегральные уравнения такого типа, в которых одна неизвестная функция $P(\mu)$ удовлетворяет различным уравнениям в двух разных областях изменения параметра x, называются «дуальными» [7].

Методы Копсона, Швингера и др., о котором упоминалось в § 11.1, отличается от приведенного выше только тем, что в нем используется одно интег-

ральное уравнение. Хотя их метод здесь и не применяется, следует отметить его связь с предлагаемым методом. Например, в случае E-поляризации решение (18), полученное с использованием преобразования Фурье, можно записать в виде

$$P(\mu) = -\frac{k}{c}\int_M J_z(\xi)\exp(-ik\mu\xi)\,d\xi, \tag{21}$$

конечно, в согласии с уравнением (8) и с тем, что $I_z(\xi)=0$ на A. Подставляя это значение $P(\mu)$ в (17) и интегрируя по μ, получим интегральное уравнение

$$\frac{k}{c}\int_M J_z(\xi)\,H_0^{(1)}(k|x-\xi|)\,d\xi = E_z^{(i)} \quad \text{на} \quad M, \tag{22}$$

содержащее функцию Хенкеля $H_0^{(1)}$ первого рода нулевого порядка; оно должно быть решено относительно $J_z(\xi)$. Очевидно, левую часть (22) можно вывести непосредственно из соотношения для дифрагировавшего поля, выраженного через плотность индуцированного тока.

§ 11.5. Двумерная дифракция плоской волны на полуплоскости

11.5.1. Решение дуальных интегральных уравнений для случая E-поляризации. На нескольких последующих страницах дифракция плоской волны на полубесконечном плоском листке рассматривается строго с помощью простого точного решения соответствующих дуальных интегральных уравнений.

Рассмотрим сначала E-поляризованную плоскую волну

$$E_z^{(i)} = \exp[-ikr\cos(\theta-\alpha_0)], \tag{1}$$

падающую на идеально проводящую полуплоскость $y=0$, $x>0$, где для удобства предполагается, что α_0 вещественно и $0<\alpha_0<\pi$ (рис. 11.6). Уравнения (11.4.17) и (11.4.18) запишутся тогда в виде

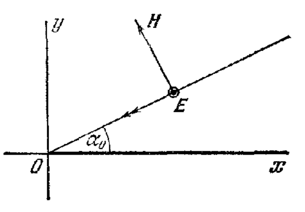


Рис. 11.6. Плоская волна, падающая на идеально проводящую полуплоскость.

$$\int_{-\infty}^{\infty}\frac{P(\mu)}{\sqrt{1-\mu^2}}\exp(ikx\mu)\,d\mu = -\exp(-ikx\mu_0) \quad \text{при} \quad x>0, \tag{2}$$

$$\int_{-\infty}^{\infty}P(\mu)\exp(ikx\mu)\,d\mu = 0 \quad \text{при} \quad x<0, \tag{3}$$

где $\mu_0=\cos\alpha_0$. Воспользуемся для их решения обычной техникой контурного интегрирования.

В интеграле в левой части уравнения (3) x отрицательно. Следовательно, согласно лемме Иордана [20] и при условии, что $P(\mu)\to 0$, когда $|\mu|\to\infty$ и $0\geqslant\arg\mu\geqslant-\pi$, мы можем замкнуть путь интегрирования бесконечной полуокружностью *ниже* вещественной оси, не внося никакого дополнительного вклада в интеграл. Таким образом, дальше необходимо только потребовать отсутствия у $P(\mu)$ сингулярностей в полуплоскости ниже пути интегрирования, чтобы уравнение (3) удовлетворялось, так как в этом случае интеграл берется по замкнутому контуру, внутри которого подынтегральное выражение регулярно.

Аналогично в интеграле левой части уравнения (2) x положительно и можно замкнуть путь интегрирования бесконечной полуокружностью *выше* вещественной оси, не внося дополнительного вклада в интеграл при условии,

что $P(\mu)/\sqrt{1-\mu^2}\to 0$, когда $|\mu|\to\infty$ и $\pi\geqslant\arg\mu\geqslant 0$. Тогда, если $U(\mu)$ — произвольная функция, свободная от сингулярностей в полуплоскости выше пути интегрирования, с соответствующим поведением там при $|\mu|\to\infty$, то, очевидно, (2) удовлетворяется при

$$\frac{P(\mu)}{\sqrt{1-\mu^2}}=-\frac{1}{2\pi i}\frac{U(\mu)}{U(-\mu_0)}\frac{1}{(\mu+\mu_0)},\qquad(4)$$

если путь интегрирования огибает полюс $\mu=-\mu_0$ снизу, как показано схематически на рис. 11.7. Действительно, единственная возможная сингулярность функции в правой части уравнения (4) — это полюс при $\mu=-\mu_0$ с вычетом, равным $-1/(2\pi i)$, а по теореме вычетов Коши это как раз дает в интеграле (2) член — $\exp(-ikx\mu_0)$.

Рис. 11.7. Путь интегрирования в комплексной μ-плоскости.

Переписав теперь (4) в виде

$$\frac{P(\mu)}{\sqrt{1-\mu}}(\mu+\mu_0)=-\frac{1}{2\pi i}\frac{U(\mu)}{U(-\mu_0)}\sqrt{1+\mu},\qquad(5)$$

можно доказать, что правая и левая части (5) постоянны. Левая часть свободна от сингулярностей в полуплоскости ниже пути интегрирования и по абсолютной величине возрастает там до бесконечности, тогда как правая часть ведет себя таким же образом в полуплоскости выше пути интегрирования. Следовательно, функция, с которой совпадают обе части, свободна от особенностей и абсолютная величина ее возрастает до бесконечности на всей комплексной μ-плоскости, значит, такая функция должна быть полиномом, а так как для некоторых значений $\arg\mu$ величина $P(\mu)\to 0$ при $|\mu|\to\infty$, этот полином может содержать только постоянный член.

Значение его сразу же можно найти, если положить $\mu=-\mu_0$ в правой части (5); тогда

$$P(\mu)=\frac{i}{2\pi}\frac{\sqrt{1-\mu_0}\,\sqrt{1-\mu}}{\mu+\mu_0},\qquad(6)$$

или

$$P(\cos\alpha)=\frac{i}{\pi}\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\alpha\right)}{\cos\alpha+\cos\alpha_0}.\qquad(7)$$

О значении симметрии (7) относительно α и α_0 говорится в конце п. 11.7.1.

Используя величину $P(\cos\alpha)$, определяемую (7), найдем из соотношений (11.4.10), (11.4.11) и (11.4.12) компоненты дифрагировавшего поля и, следовательно, для полного поля получим

$$E_z=\exp\left[-ikr\cos(\theta-\alpha_0)\right]-$$
$$-\frac{1}{i\pi}\int\limits_C\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\alpha\right)}{\cos\alpha+\cos\alpha_0}\exp\left[ikr\cos(\theta\mp\alpha)\right]d\alpha.\qquad(8)$$

Верхний знак берется при $y>0$, нижний — при $y<0$. Теперь остается придать полученному решению более удобную форму.

11.5.2. Выражение решения через интегралы Френеля. Если kr велико, т. е. если расстояния от начала координат порядка длины волны или больше нее, можно попытаться вычислить интегралы общего вида

$$\int P(\cos\alpha)\exp\left[ikr\cos(\theta-\alpha)\right]d\alpha\qquad(9)$$

методом наибыстрейшего спуска (см. приложение 3). Предварительная процедура в этом методе заключается в замене пути интегрирования (сейчас в подын-

тегральном выражении допускается существование любых сингулярностей) на путь наибыстрейшего спуска, например $S(\theta)$, проходящий через седловую точку при $\alpha=\theta$. Путь $S(\theta)$ показан на рис. 11.8. Вдоль него новая переменная

$$\tau=\sqrt{2}\exp\left(\frac{1}{4}i\pi\right)\sin\frac{1}{2}(\alpha-\theta) \tag{10}$$

пробегает все вещественные значения от $-\infty$ до $+\infty$. Тогда интеграл (9) принимает вид

$$\sqrt{2}\exp\left(-\frac{1}{4}i\pi\right)\exp(ikr)\int_{-\infty}^{\infty}\frac{P(\cos\alpha)}{\sqrt{1+\frac{1}{2}i\tau^2}}\exp(-ir\tau^2)\,d\tau. \tag{11}$$

Отсюда можно получить асимптотическое приближение для $kr\gg 1$.

Применение такого метода к специальному интегралу в (8) позволяет фактически без приближений получить его выражение через интегралы Френеля. Сейчас это будет показано.

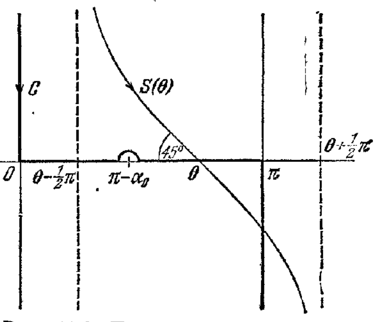


Рис. 11.8. Путь наибыстрейшего спуска $S(\theta)$ в комплексной α-плоскости.

Рассмотрим сначала случай $0<\theta<\pi$. Так как (7) можно представить в виде

$$P(\cos\alpha)=\frac{1}{4\pi i}\left\{\sec\frac{1}{2}(\alpha-\alpha_0)-\sec\frac{1}{2}(\alpha+\alpha_0)\right\}, \tag{12}$$

то достаточно вычислить

$$\int_{S(\theta)}\sec\frac{1}{2}(\alpha-\alpha_0)\exp(ikr\cos(\theta-\alpha))\,d\alpha, \tag{13}$$

поскольку вклад от $\sec\frac{1}{2}(\alpha+\alpha_0)$ можно затем найти, изменяя знак α_0. Простое преобразование (13) дает

$$\int_{S(0)}\sec\frac{1}{2}(\alpha-\alpha_0+\theta)\exp(ikr\cos\alpha)\,d\alpha=$$

$$=\frac{1}{2}\int_{S(0)}\left\{\sec\frac{1}{2}(\alpha-\alpha_0+\theta)+\sec\frac{1}{2}(\alpha+\alpha_0-\theta)\right\}\exp(ikr\cos\alpha)\,d\alpha=$$

$$=\frac{1}{2}\int_{S(0)}\frac{\cos\frac{1}{2}(\alpha_0-\theta)\cos\left(\frac{1}{2}\alpha_0\right)}{\cos\alpha+\cos(\alpha_0-\theta)}\exp(ikr\cos\alpha)\,d\alpha; \tag{14}$$

применив подстановку

$$\tau=\sqrt{2}\exp\left(\frac{1}{4}i\pi\right)\sin\left(\frac{1}{2}\alpha\right),$$

получим

$$-2\exp\left(\frac{1}{4}i\pi\right)\exp(ikr)\,\eta\int_{-\infty}^{\infty}\frac{\exp(-kr\tau^2)}{\tau^2-i\eta^2}\,d\tau, \tag{15}$$

где

$$\eta=\sqrt{2}\cos\frac{1}{2}(\theta-\alpha_0).$$

Но

$$\int_{-\infty}^{\infty}\exp(-\xi\tau^2)\,d\tau=\sqrt{\frac{\pi}{\xi}},$$

откуда, умножая на $\exp(i\eta^2\xi)$ и интегрируя по ξ от kr до бесконечности, получим

$$\exp(ikr\eta^2)\int_{-\infty}^{\infty}\frac{\exp(-kr\tau^2)}{\tau^2-i\eta^2}\,d\tau=\sqrt{\pi}\int_{kr}^{\infty}\frac{\exp(i\eta^2\xi)}{\sqrt{\xi}}\,d\xi=\frac{2\sqrt{\pi}}{|\eta|}\int_{|\eta|\sqrt{kr}}^{\infty}\exp(i\mu^2)\,d\mu. \tag{16}$$

Обозначая через

$$F(a)=\int_a^{\infty}\exp(i\mu^2)\,d\mu \tag{17}$$

одну из форм комплексного интеграла Френеля *), находим

$$\eta\int_{-\infty}^{\infty}\frac{\exp(-kr\tau^2)}{\tau^2-i\eta^2}\,d\tau=\pm 2\sqrt{\pi}\exp(-ikr\eta^2)F\{\pm\eta\sqrt{kr}\}. \tag{18}$$

Верхний знак берется при $\eta>0$, нижний — при $\eta<0$.

Объединяя эти результаты, имеем окончательно для $y\geqslant 0$

$$-\frac{1}{i\pi}\int_{S(\theta)}\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\alpha\right)}{\cos\alpha+\cos\alpha_0}\exp[ikr\cos(\theta-\alpha)]\,d\alpha=$$

$$=-\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\left\{\exp[-ikr\cos(\theta-\alpha_0)]F\left[\sqrt{2kr}\cos\frac{1}{2}(\theta-\alpha_0)\right]\mp\right.$$
$$\left.\mp\exp[-ikr\cos(\theta+\alpha_0)]F\left[\pm\sqrt{2kr}\cos\frac{1}{2}(\theta+\alpha_0)\right]\right\}, \tag{19}$$

где верхний знак берется для $\theta+\alpha_0<\pi$, а нижний — для $\theta+\alpha_0>\pi$.

Для того чтобы получить из (8) полное поле, остается только учесть простой полюс при $\alpha=\pi-\alpha_0$. Поскольку $0\leqslant\theta\leqslant\pi$, легко показать, что при замене пути C на путь $S(\theta)$ части на бесконечности не вносят вклада в поле, и из рис. 11.8 ясно, что полюс захватывается тогда и только тогда, когда $\pi-\alpha_0>\theta$. Его вклад, полученный по теореме вычетов, равен

$$\exp[-ikr\cos(\theta+\alpha_0)]. \tag{20}$$

Другими словами, он представляет отраженную волну *геометрической оптики*, разрыв которой при переходе через $\theta=\pi-\alpha_0$ точно уравновешивается разрывом в *дифрагированном* поле (19). В самом деле, применяя соотношение

$$F(a)+F(-a)=\sqrt{\pi}\exp\left(\frac{1}{4}i\pi\right), \tag{21}$$

мы можем записать полное поле (8) в виде

$$E_z=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\left\{\exp[-ikr\cos(\theta-\alpha_0)]F\left[-\sqrt{2kr}\cos\frac{1}{2}(\theta-\alpha_0)\right]-\right.$$
$$\left.-\exp[-ikr\cos(\theta+\alpha_0)]F\left[-\sqrt{2kr}\cos\frac{1}{2}(\theta+\alpha_0)\right]\right\}, \tag{22}$$

что совпадает с известным результатом Зоммерфельда.

Когда $y<0$, вычисляемый интеграл равен

$$\int_C\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\alpha\right)}{\cos\alpha+\cos\alpha_0}\exp[ikr\cos(\theta+\alpha)]\,d\alpha.$$

---

*) Эта форма интеграла Френеля здесь более удобна, чем форма (8.7.12) Следует отметить изменение пределов интегрирования.

Соответствующий путь наибыстрейшего спуска теперь имеет вид $S(2\pi-\theta)$, и захват полюса при $\alpha=\pi-\alpha_0$, который имеет место только при $\theta>\pi+\alpha_0$, приводит к выражению, соответствующему падающей волне со знаком минус. Полное поле опять определяется выражением (22).

Соответствующие соотношения для компонент **H** получаются просто дифференцированием, как показано в п. 11.4.1. Представляют интерес как декартовы компоненты H_x и H_y, так и полярные компоненты H_r и H_θ. Ввиду того, что (22) выражено через r и θ, удобно сначала получить H_r и H_θ из уравнений Максвелла

$$H_r=\frac{1}{ikr}\frac{\partial E_z}{\partial\theta},\qquad H_\theta=-\frac{1}{ik}\frac{\partial E_z}{\partial r},\tag{23}$$

а затем найти H_x и H_y из соотношений

$$H_x=\cos\theta H_r-\sin\theta H_\theta,\qquad H_y=\sin\theta H_r+\cos\theta H_\theta.$$

Для получения более компактных результатов введем следующие обозначения:

$$u=-\sqrt{2kr}\cos\frac{1}{2}(\theta-\alpha_0),\qquad v=-\sqrt{2kr}\cos\frac{1}{2}(\theta+\alpha_0),\tag{24}$$

$$G(a)=\exp(-ia^2)F(a).\tag{25}$$

Заметим, что

$$\frac{dG(a)}{da}=-1-2iaG(a).$$

Выражение (22) тогда принимает вид

$$E_z=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\{G(u)-G(v)\};\tag{26}$$

отсюда следует, что

$$\left.\begin{aligned}H_r=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{&\sin(\theta-\alpha_0)G(u)-\\&-\sin(\theta+\alpha_0)G(v)-i\sqrt{\frac{2}{kr}}\sin\left(\frac{1}{2}\alpha_0\right)\cos\left(\frac{1}{2}\theta\right)\Big\},\\H_\theta=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{&\cos(\theta-\alpha_0)G(u)-\\&-\cos(\theta+\alpha_0)G(v)+i\sqrt{\frac{2}{kr}}\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)\Big\},\end{aligned}\right\}\tag{27}$$

$$\left.\begin{aligned}H_x=-\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{&\sin\alpha_0\,[G(u)+G(v)]+\\&+i\sqrt{\frac{2}{kr}}\sin\left(\frac{1}{2}\alpha_0\right)\cos\left(\frac{1}{2}\theta\right)\Big\},\\H_y=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{&\cos\alpha_0\,[G(u)-G(v)]-\\&-i\sqrt{\frac{2}{kr}}\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)\Big\}.\end{aligned}\right\}\tag{28}$$

11.5.3. Характер решения. Теперь исследуем подробнее характер результатов, полученных в п. 11.5.2. Из приведенного там вывода очевидно (и это можно прямо проверить), что $G(u)\exp(ikr)$ есть решение двумерного волнового уравнения при любом значении α_0; здесь интересно отметить, что оно имеет

периодичность 4π и, следовательно, $G(u) - G(v)$ исчезает при $\theta = 0$ и $\theta = 2\pi$, т. е. на обеих поверхностях экрана, но не исчезает при $\theta = \pi$. Зоммерфельд пришел к своему решению (22), пытаясь найти соответствующее решение волнового уравнения с периодом 4π и объединяя это решение с его «изображением» *). Кстати, из (28) вытекает, что выражения

$$\cos\left(\frac{1}{2}\theta\right)\frac{\exp(ikr)}{\sqrt{kr}}, \quad \sin\left(\frac{1}{2}\theta\right)\frac{\exp(ikr)}{\sqrt{kr}}$$

также служат решениями двумерного волнового уравнения, и этот результат хорошо известен.

Следует исследовать еще поведение (26) при $r \to \infty$. Это очевидно и будет в дальнейшем главной темой, обсуждаемой в данном разделе.

Привлекательной чертой задачи с полуплоскостью является то, что в любой точке поле можно найти, пользуясь таблицей интегралов Френеля **). Кроме того, в двух особо интересных случаях, а именно при $kr \gg 1$ и $kr \ll 1$ пригодны простые приближения к интегралам Френеля (см. п. 8.4.2). Первое условие, без сомнения, всегда выполняется в оптических экспериментах, где точка наблюдения находится, вероятно, на расстоянии миллионов длин волн от дифракционного края; второе условие выполняется при изучении поведения поля вблизи острого края. Этот случай можно изучать на сантиметровых радиоволнах (см. п. 11.5.6).

Случай $kr \gg 1$. В данном случае $|u|$ и $|v|$ велики по сравнению с единицей, за исключением значений θ, достаточно близких к $\pi + \alpha_0$ и $\pi - \alpha_0$ соответственно. Для определенности введем пять областей, как показано на рис. 11.9.

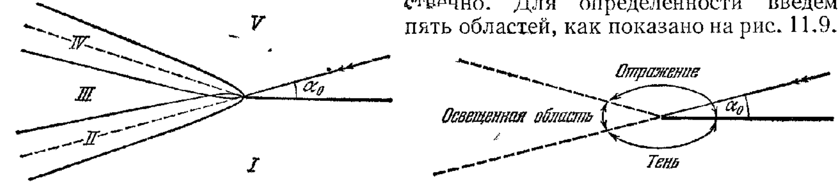


Рис 11.9. Пять областей, в которых описывается поведение поля при дифракции плоской волны на идеально проводящей полуплоскости.

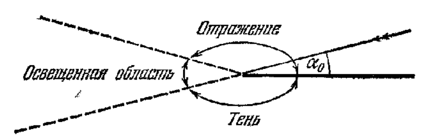


Рис 11.10 Три области геометрической оптики при дифракции плоской волны на идеально проводящей полуплоскости.

За уравнения кривых, ограничивающих области II и IV, примем соответственно $u^2 = 1$ и $v^2 = 1$. Следовательно, эти кривые имеют вид парабол с фокусами в начале координат и осями $\theta = \pi + \alpha_0$ и $\theta = \pi - \alpha_0$. В глубине области II (т. е. внутри параболы $u^2 = \varepsilon$, где $\varepsilon \ll 1$) $|u| \ll 1$; далеко от области II (т. е. вне параболы $u^2 = \gamma$, где $\gamma \gg 1$) $|u| \gg 1$. Аналогичные соотношения получим и для $|v|$ и области IV. Кроме того, при $0 < \theta < \pi - \alpha_0$ как u, так и v отрицательно; при $\pi - \alpha_0 < \theta < \pi + \alpha_0$ u отрицательно, но v положительно; при $\pi + \alpha_0 < \theta < 2\pi$ как u, так и v положительно.

Области I, III и V, очевидно, тесно связаны с теми областями, которые появились бы в рамках геометрической оптики, где свет распространяется по прямым линиям. Перечислим их: область *тени* позади экрана, где поля нет совсем, *освещенная* область, где существует только падающая плоская волна и область *отражения*, где одновременно присутствуют падающая плоская волна и отраженная плоская волна, соответствующая отражению от *бесконечного* экрана (рис. 11.10). Вообще говоря, в областях II и IV осуществляется плавный переход от точного решения в одной области геометрической оптики к решению в соседней области. Прежде чем показать это более подробно, мы должны за-

*) Хорошее представление об этом подходе можно получить из гл. 4 книги [16].
**) По-видимому, лучше всего подходят для этой цели таблицы, приведенные в [21].

держаться, для того чтобы вывести асимптотическое приближение для интеграла Френеля при больших значениях аргумента.

Если а положительно, можно написать

$$G(a)=\exp(-ia^2)\int_a^\infty \frac{d[\exp(i\mu^2)]}{2i\mu};$$

дважды интегрируя по частям, получим

$$G(a)=\frac{i}{2a}+\frac{1}{4a^3}+\frac{3}{4}\exp(-ia^2)\int_a^\infty \frac{\exp(i\mu^2)}{\mu^4}d\mu. \tag{29}$$

Продолжая этот процесс и дальше, можно получить полное асимптотическое разложение для $G(a)$, но сейчас нам важно заметить только, что модуль интеграла в последнем члене (29) меньше чем

$$\int_a^\infty \frac{d\mu}{\mu^4}=\frac{1}{3a^3}. \tag{30}$$

Отсюда имеем

$$G(a)=\frac{i}{2a}+O\left(\frac{1}{a^3}\right). \tag{31}$$

Здесь интересно отметить, что этот результат можно было получить и общим методом, описанным в п. 11.5.2, который в данном случае сводился бы к разложению множителя $(\tau^2-i\eta^2)^{-1}$ в подынтегральном выражении (18) в ряд по степеням τ и к интегрированию каждого его члена.

Если а отрицательно, то левая часть (30) расходится, но этот случай легко рассмотреть, используя (21) совместно с результатом, полученным для положительных значений аргумента. Таким образом,

$$G(a)=\sqrt{\pi}\exp\left(\frac{1}{4}i\pi\right)\exp(-ia)+\frac{i}{2a}+O\left(\frac{1}{a^3}\right). \tag{32}$$

То обстоятельство, что асимптотическое приближение (31) для положительного a и (32) для отрицательного a различны, есть частный случай явления Стокса [22]. Напишем теперь

$$E_z=E_z^{(g)}+E_z^{(d)},$$

где $E_z^{(g)}$ — поле геометрической оптики, определяемое как

$$E_z^{(g)}=\begin{cases}\exp[-ikr\cos(\theta-\alpha_0)]-\exp[-ikr\cos(\theta+\alpha_0)] & \text{для} \quad 0\leqslant\theta<\pi-\alpha_0,\\ \exp[-ikr\cos(\theta-\alpha_0)] & \text{для} \quad \pi-\alpha_0<\theta<\pi+\alpha_0,\\ 0 & \text{для} \quad \pi+\alpha_0<\theta\leqslant 2\pi,\end{cases} \tag{33}$$

а $E_z^{(d)}$ — поле *дифракции*, представляющее собой просто то поле, которое следует добавить к полю геометрической оптики, чтобы получить полное поле. Теперь при $kr\gg 1$ находим из (26), используя (31) и (32),

$$E_z^{(d)}\approx\sqrt{\frac{2}{\pi}}\exp\left(\frac{1}{4}i\pi\right)\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)}{\cos\theta+\cos\alpha_0}\frac{\exp(-ikr)}{\sqrt{kr}} \tag{34}$$

в точках, не слишком близких к областям II и IV (в смысле, указанном выше). Легко видеть либо из (23), либо из (27), что в том же приближении, как и в (34), компоненты $\mathbf{H}^d$ равны $H_\theta^{(d)}=-E_z^{(d)}$ и $H_r^{(d)}=0$. Очевидно, (34) означает, что поле дифракции ведет себя так, как будто оно порождается линейным источником, расположенным вдоль дифракционного края, с «полярной диаграммой», изменяющейся с углом, как было определено выше. Это находится в

согласии с экспериментальным фактом, что дифракционный край кажется светящимся, если он рассматривается из области тени.

Когда $\cos\theta + \cos\alpha_0$ приближается к нулю, соотношение (34) становится неверным, и нужно обратиться к точному решению. Так как

$$G(0) = \int_0^\infty \exp(i\mu^2)\,d\mu = \frac{1}{2}\sqrt{\pi}\exp\left(\frac{1}{4}i\pi\right), \tag{35}$$

то из (26) следует, что при $\theta = \pi + \alpha_0$

$$E_z = \frac{1}{2}\exp(ikr) + O\left\{\frac{1}{\sqrt{kr}}\right\}, \tag{36}$$

а при $\theta = \pi - \alpha_0$ —

$$E_z = \exp[ikr\cos(2\alpha_0)] - \frac{1}{2}\exp(ikr) + O\left\{\frac{1}{\sqrt{kr}}\right\}. \tag{37}$$

Следовательно, вблизи $\theta = \pi + \alpha_0$ и $\theta = \pi - \alpha_0$ поле дифракции имеет тот же порядок, что и падающее поле. В частности, в бесконечности переход между полями геометрической оптики в соседних областях происходит через их среднее арифметическое.

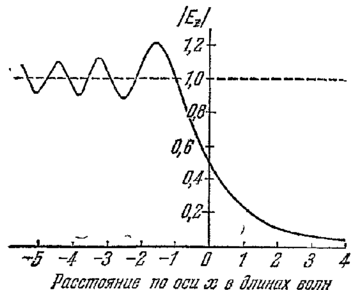


Рис. 11.11. Дифракция нормально падающей E-поляризованной плоской волны единичной амплитуды на идеально проводящей полуплоскости. Изменение $|E_z|$ в зависимости от x на расстоянии трех длин волн позади экрана.

Интерференция поля геометрической оптики с полем дифракции в областях, где они сравнимы, вызывает возникновение полос. Они видны на рис. 11.11, который будет обсуждаться ниже.

Случай $kr \ll 1$. В данном случае $|u|$ и $|v|$ малы по сравнению с единицей, и полезно использовать разложение в ряд интеграла Френеля. Напишем

$$F(a) = \int_0^\infty \exp(i\mu^2)\,d\mu - \int_0^a \exp(i\mu^2)\,d\mu;$$

разлагая экспоненту в подынтегральном выражении второго интеграла, получим

$$F(a) = \frac{1}{2}\sqrt{\pi}\exp\left(\frac{1}{4}i\pi\right) - a + O(a^3). \tag{38}$$

Таким образом, из (26) и (28) находим, пренебрегая членами с kr в степени, большей половины,

$$\left.\begin{aligned}
E_z &= 2\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\sqrt{kr}\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right),\\
H_x &= -\sin\alpha_0 - \sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\sin\left(\frac{1}{2}\alpha_0\right)\cos\left(\frac{1}{2}\theta\right)\times\\
&\qquad\times\left\{\frac{i}{\sqrt{kr}} + (1+2\cos\alpha_0)\sqrt{kr}\right\},\\
H_y &= \sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)\times\\
&\qquad\times\left\{-\frac{i}{\sqrt{kr}} + (1+2\cos\alpha_0)\sqrt{kr}\right\}.
\end{aligned}\right\} \tag{39}$$

Необходимо отметить, что E_z конечно и непрерывно при $r=0$, но H_x и H_y расходятся как $r^{-1/2}$; исключение составляют случаи $\theta=\pi$, когда $H_x = -\sin\alpha_0\exp(ikr\cos\alpha_0)$, а также $\theta=0$, $\theta=2\pi$, когда $H_y=0$. Такое пове-

дение функции, необычное для физических задач, является, конечно, следствием идеализированного представления о бесконечно остром крае. В этом случае существование сингулярностей у компонент поля должно быть учтено при формулировке любой теоремы, относящейся к единственности решения (см. ниже, § 11.9).

Мы закончим наше исследование решения рассмотрением плотности тока, индуцированного в дифракционном экране. Она равна произведению $-c/4\pi$ на разность значений H_x при $\theta=0$ и $\theta=2\pi$. Таким образом, из (28) находим

$$\frac{2\pi}{c} J_z = \sin\alpha_0 \exp(-ikx\cos\alpha_0)\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikx)\times$$
$$\times\left\{2\sin\alpha_0 G\left[\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\right]-i\sqrt{\frac{2}{kx}}\sin\left(\frac{1}{2}\alpha_0\right)\right\}. \quad (40)$$

При $\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\gg 1$ последнее соотношение принимает вид

$$J_z = \frac{c}{2\pi}\sin\alpha_0\exp(-ikx\cos\alpha_0)+O\{(kx)^{-3/2}\}. \quad (41)$$

Полученный результат интересен тем, что он показывает ту степень приближения, с которой величину плотности тока можно считать совпадающей с величиной, получающейся в рамках геометрической оптики; это обычный метод в задачах, которые невозможно решить точно. Очевидно, что введенное выше допущение $\left(\text{т. е. } \sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\gg 1\right)$ разумно только для значений α_0, не близких к π; оно имеет смысл потому, что при $x\to\infty$ «остаточный» член в (41) стремится к нулю, как $x^{-3/2}$, а не как $x^{-1/2}$.

Вместе с тем при $\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\ll 1$

$$J_z = \frac{c}{\pi\sqrt{2\pi}}\exp\left(-\frac{1}{4}i\pi\right)\sin\left(\frac{1}{2}\alpha_0\right)\left\{\frac{i}{\sqrt{kx}}+4\cos^2\left(\frac{1}{2}\alpha_0\right)\sqrt{kx}\right\}\exp(ikx) \quad (42)$$

и расходится на дифракционном крае.

11.5.4. Решение для H-поляризации. Случай H-поляризации, когда падающее поле определяется выражением

$$H_z^{(i)} = \exp[-ikr\cos(\theta-\alpha_0)], \quad (43)$$

можно рассматривать таким же способом, как и случай E-поляризации; действительно, анализ оказывается практически идентичным. Вместе с тем первый случай можно вывести из последнего с помощью точной электромагнитной формы принципа Бабине, полученного в § 11.3, так как дополнительным экраном к полуплоскости служит также полуплоскость. Очевидно, что полное поле определяется выражением

$$H_z = \frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\left\{\exp[-ikr\cos(\theta-\alpha_0)]\,F\left[-\sqrt{2kr}\cos\frac{1}{2}(\theta-\alpha_0)\right]+\right.$$
$$\left.+\exp[-ikr\cos(\theta+\alpha_0)]\,F\left[-\sqrt{2kr}\cos\frac{1}{2}(\theta+\alpha_0)\right]\right\}. \quad (44)$$

Оно отличается от соответствующего выражения для E_z в случае поля с E-поляризацией (22) только знаком второго члена.

Применяя обозначения, введенные в п. 11.5.2, получим для компонент поля, не равных нулю, выражения

$$\left.\begin{aligned} H_z &= \frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\{G(u)+G(v)\},\\ E_x &= \frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{\sin\alpha_0\,[G(u)-G(v)]-\\ &\qquad -i\sqrt{\frac{2}{kr}}\cos\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)\Big\},\\ E_y &= -\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikr)\Big\{\cos\alpha_0\,[G(u)+G(v)]-\\ &\qquad -i\sqrt{\frac{2}{kr}}\cos\left(\frac{1}{2}\alpha_0\right)\cos\left(\frac{1}{2}\theta\right)\Big\}. \end{aligned}\right\} \quad (45)$$

Несомненно, E_x исчезает при $\theta=0$ и $\theta=2\pi$, и для интерпретации поведения поля при $kr \gg 1$ снова можно воспользоваться полем дифракции (которое должно создаваться линейным источником, расположенным вдоль дифракционного края) для точек, достаточно удаленных от $\theta=\pi-\alpha_0$ и $\theta=\pi+\alpha_0$. При $r\to 0$ H_z остается конечным и непрерывным, тогда как E_x и E_y расходятся как $r^{-1/2}$; исключением являются $E_x=0$ при $\theta=0$, 2π и $E_y=-\cos\alpha_0\exp(ikr\cos\alpha_0)$ при $\theta=\pi$.

Плотность тока определяется выражением

$$\frac{2\pi}{c}J_x=\exp\left[-ikx\cos\alpha_0\right]-\frac{2\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp(ikx)\,G\left(\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\right). \quad (46)$$

При $\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\gg 1$ имеем

$$\frac{2\pi}{c}J_x=\exp(-ikx\cos\alpha_0)-\frac{\exp\left(\frac{1}{4}i\pi\right)}{\sqrt{2\pi}}\sec\left(\frac{1}{2}\alpha_0\right)\frac{\exp(ikx)}{\sqrt{kx}}; \quad (47)$$

это выражение приближается менее быстро к плотности тока, определяемой геометрической оптикой, чем в случае E-поляризации. При $\sqrt{2kx}\cos\left(\frac{1}{2}\alpha_0\right)\ll 1$ имеем

$$J_x=\frac{c}{\pi}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\cos\left(\frac{1}{2}\alpha_0\right)\sqrt{kx}\exp(ikx). \quad (48)$$

Это выражение обращается в нуль при $x=0$, так что, как следовало ожидать, на самом крае не существует тока, нормального к краю.

11.5.5. Некоторые численные расчеты. Типичная теоретическая кривая, полученная из (26), приведена на рис. 11.11. Она построена для случая нормального падения E-поляризованной плоской волны единичной амплитуды и представляет собой график изменения амплитуды E_z в зависимости от x на расстоянии трех длин волн позади экрана ($ky=-6\pi$). На рисунке отчетливо видны дифракционные полосы в освещенной области и постепенное ослабление поля в глубине области геометрической тени.

Несколько интересных расчетов было выполнено Браунбеком и Лаукиеном [23]. Они построили кривые равных амплитуд (рис. 11.12) и кривые равных фаз (рис. 11.13) компоненты H_z для нормально падающей H-поляризованной плоской волны единичной амплитуды на расстоянии от дифракционного

края, приблизительно равной одной длине волны. Они рассчитали также линии среднего потока энергии (рис. 11.14), ортогональные к линиям равных фаз.

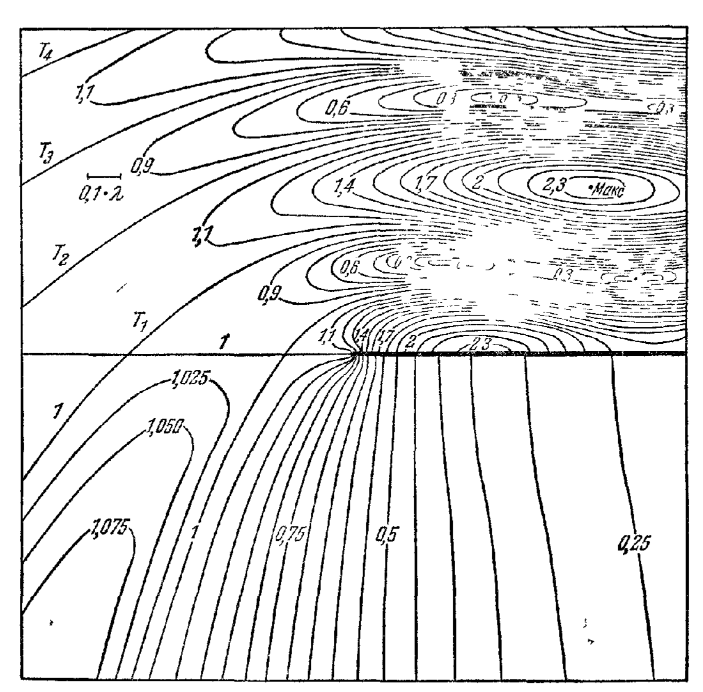


Рис. 11.12. Линии равных амплитуд компоненты H_z при дифракции нормально падающей H-поляризованной плоской волны на идеально проводящей полуплоскости [23].

Амплитуда падающей волны принята за единицу.

Легко показать, что найденные результаты справедливы для *любого* двумерного поля с H-поляризацией. Напишем $H_z = he^{i\varphi}$, где h и φ вещественны; тогда, используя соотношения

$$E_x = -\frac{1}{ik}\frac{\partial H_z}{\partial y}, \qquad E_y = \frac{1}{ik}\frac{\partial H_z}{\partial x},$$

получим, что усредненный вектор Пойнтинга (см. (1.4.56))

$$\frac{c}{8\pi}\,\mathrm{Re}\,(\mathbf{E}\times\mathbf{H}^*) = \frac{c}{8\pi}\,\mathrm{Re}\,(E_yH_z^*,\ -E_xH_z^*,\ 0),$$

равный

$$\frac{c}{8\pi}\frac{h^2}{k}\left(\frac{\partial\varphi}{\partial x},\ \frac{\partial\varphi}{\partial y},\ 0\right), \tag{49}$$

перпендикулярен к поверхности $\varphi = \mathrm{const}$. Аналогичный результат получается и для поля с E-поляризацией.

11.5.6. Сравнение с приближенной теорией и с экспериментальными результатами. Для точек, находящихся на большом расстоянии от дифракционного края в освещенной части области II (см. рис. 11.9), там, где видны полосы, можно пренебречь вторым членом в каждом из решений (22) и (44). Таким образом, интенсивность как при E-поляризации, так и при H-поляризации, а

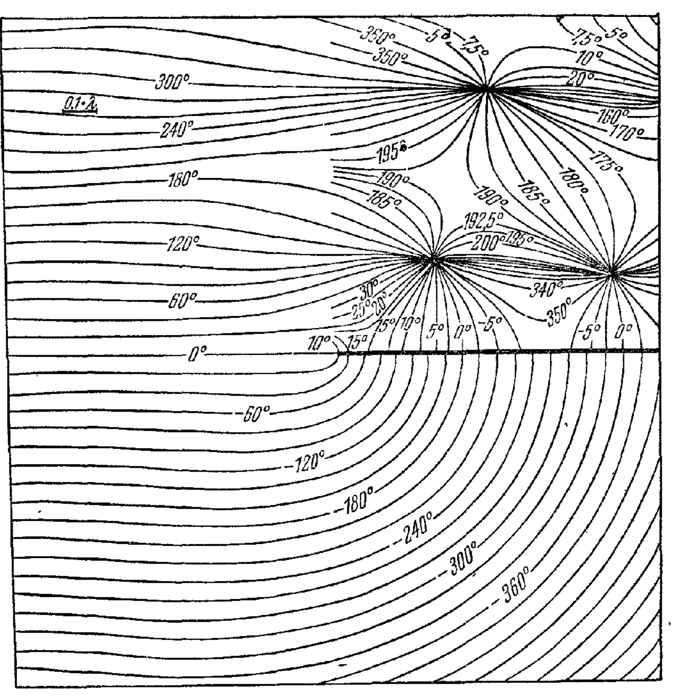


Рис 11.13. Линии равных фаз компоненты H_z при дифракции нормально падающей H-поляризованной плоской волны на идеально проводящей полуплоскости [23].

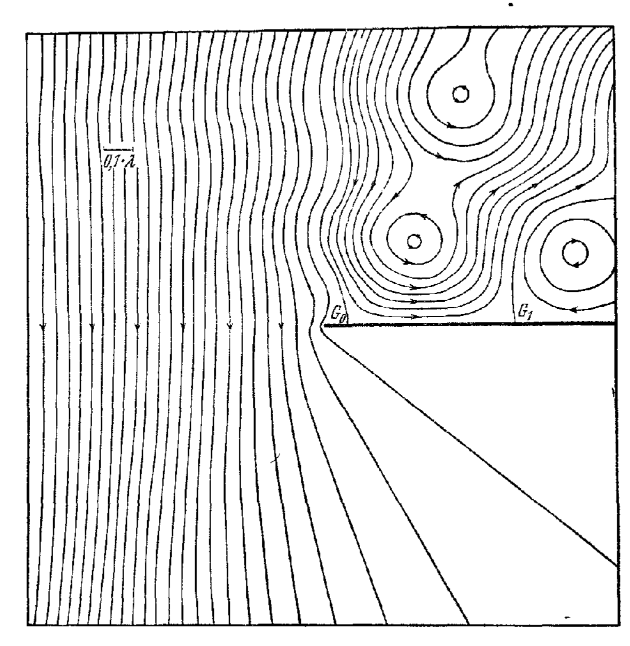


Рис. 11.14. Линии среднего потока энергии при дифракции нормально падающей H-поляризованной плоской волны на идеально проводящей полуплоскости [23].

следовательно и интенсивность неполяризованного света, запишется в виде

$$\frac{1}{2}\left\{\frac{1}{2}+\mathscr{C}\left[2\sqrt{\frac{2r}{\lambda}}\cos\frac{1}{2}(\theta-\theta_0)\right]\right\}^2+\frac{1}{2}\left\{\frac{1}{2}+\mathscr{S}\left[2\sqrt{\frac{2r}{\lambda}}\cos\frac{1}{2}(\theta-\theta_0)\right]\right\}^2, \tag{50}$$

где λ — длина волны света, а $\mathscr{C}$ и $\mathscr{S}$ — интегралы Френеля с косинусом и синусом, определенные (8.7.12). Это следует сравнить с аналогичным результатом (8.7.28) для черной полуплоскости в теории Френеля — Кирхгофа. Отсюда действительно можно сделать вывод *), что первый член точного решения для дифракции на идеально проводящей полуплоскости можно считать решением для черной полуплоскости.

Далеко от границы в области тени I поле с E-поляризацией определяется выражением (34), т. е.

$$E_z=\sqrt{\frac{2}{\pi}}\exp\left(\frac{1}{4}i\pi\right)\frac{\sin\left(\frac{1}{2}\alpha_0\right)\sin\left(\frac{1}{2}\theta\right)}{\cos\alpha_0+\cos\theta}\,\frac{\exp(ikr)}{\sqrt{kr}}. \tag{51}$$

Аналогично можно показать, что поле с H-поляризацией имеет вид

$$H_z=-\sqrt{\frac{2}{\pi}}\exp\left(\frac{1}{4}i\pi\right)\frac{\cos\left(\frac{1}{2}\alpha_0\right)\cos\left(\frac{1}{2}\theta\right)}{\cos\alpha_0+\cos\theta}\,\frac{\exp(ikr)}{\sqrt{kr}}. \tag{52}$$

Следовательно, соответствующее отношение «силы» полей равно

$$\frac{E\text{-поляризация}}{H\text{-поляризация}}=-\operatorname{tg}\left(\frac{1}{2}\alpha_0\right)\operatorname{tg}\left(\frac{1}{2}\theta\right),$$

и, значит, падающий неполяризованный свет также становится при дифракции частично поляризованным. Этот результат находится в хорошем согласии с оптическими экспериментами [15, 24].

Развитие микроволновой радиотехники открыло широкие возможности для экспериментального изучения дифракции электромагнитных волн. В частности, появилась возможность использовать дифракционный экран, значительно лучше приближающийся к идеализированной идеально проводящей полуплоскости, чем те, которые удается реализовать в оптических измерениях. Кроме того, в микроволновой области легко исследуются поля в непосредственной близости к краю дифракционного экрана. Был проведен ряд измерений, главным образом на длине волны около 3 *см*, которые показали хорошее согласие теории и эксперимента [25, 26].

§ 11.6. Трехмерная дифракция плоской волны на полуплоскости

В § 11.5 была решена задача дифракции на полуплоскости произвольной плоской волны, было введено одно только ограничение, а именно, волна должна была распространяться в направлении, нормальном к дифракционному краю. Сейчас будет показано, как простой прием позволяет распространить полученные выше результаты на полностью произвольно падающую плоскую волну.

Допустим, что падающая плоская волна характеризуется фазовым множителем

$$\exp(-ikS)=\exp\left[-ik(x\cos\alpha\cos\beta+y\sin\alpha\cos\beta+z\sin\beta)\right], \tag{1}$$

где, как и раньше, идеально проводящий экран занимает полуплоскость $y=0$, $x>0$. Углы α и β, определяющие направление распространения волны, показаны на рис. 11.15.

*) См. обзор [16].

Отметим, что соотношение (1) получается из двумерной формы, соответствующей $\beta=0$, при замене k на $k\cos\beta$ и умножении на $\exp(-ikz\sin\beta)$. В самом деле, этот метод, применимый к любому двумерному решению волнового уравнения

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+k^2V=0, \tag{2}$$

очевидно, позволяет получить решение и трехмерного волнового уравнения

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}+k^2V=0, \tag{3}$$

где z входит только через множитель $\exp(-ikz\sin\beta)$. Кроме того, если U — такое решение (3), то легко показать, что два электромагнитных поля, удовлетворяющих уравнениям Максвелла, имеют вид

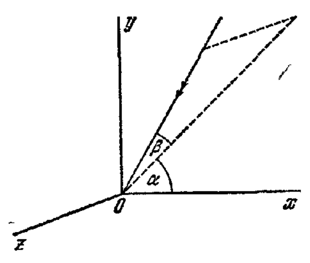


Рис. 11.15. Направление распространения падающей плоской волны.

$$\left.\begin{aligned}\mathbf{E}&=\left(-\frac{i\sin\beta}{k}\frac{\partial U}{\partial x},\ -\frac{i\sin\beta}{k}\frac{\partial U}{\partial y},\ \cos^2\beta U\right),\\ \mathbf{H}&=\left(-\frac{i}{k}\frac{\partial U}{\partial y},\ \frac{i}{k}\frac{\partial U}{\partial x},\ 0\right),\end{aligned}\right\} \tag{4}$$

и

$$\left.\begin{aligned}\mathbf{E}&=\left(\frac{i}{k}\frac{\partial U}{\partial y},\ -\frac{i}{k}\frac{\partial U}{\partial x},\ 0\right),\\ \mathbf{H}&=\left(-\frac{i\sin\beta}{k}\frac{\partial U}{\partial x},\ -\frac{i\sin\beta}{k}\frac{\partial U}{\partial y},\ \cos^2\beta U\right).\end{aligned}\right\} \tag{5}$$

При $\beta=0$ (4) определяет двумерное поле с E-поляризацией, (5) — поле с H-поляризацией.

Если мы будем считать, что U имеет вид (1), то (4) и (5) дадут соответственно две плоские волны

$$\left.\begin{aligned}\mathbf{E}&=(-\cos\alpha\sin\beta,\ -\sin\alpha\sin\beta,\ \cos\beta)\exp(-ikS),\\ \mathbf{H}&=(-\sin\alpha,\ \cos\alpha,\ 0)\exp(-ikS),\end{aligned}\right\} \tag{6}$$

и

$$\left.\begin{aligned}\mathbf{E}&=(\sin\alpha,\ -\cos\alpha,\ 0)\exp(-ikS),\\ \mathbf{H}&=(-\cos\alpha\sin\beta,\ -\sin\alpha\sin\beta,\ \cos\beta)\exp(-ikS).\end{aligned}\right\} \tag{7}$$

Здесь множитель $\cos\beta$ везде опущен. Итак, *любая* плоская волна, изменяющаяся в пространстве в соответствии с (1), определяется двумя компонентами $\mathbf{E}$ (или $\mathbf{H}$), так как третья компонента определяется уравнением $\operatorname{div}\mathbf{E}=0$ (или $\operatorname{div}\mathbf{H}=0$). Следовательно, *любую* плоскую волну можно получить соответствующей суперпозицией полей (6) и (7); поэтому в дифракционных задачах без потери общности можно ограничиться двумя случаями, относящимися к таким падающим полям.

Теперь должно быть ясно, что решение дифракционной задачи определяется соотношениями (4), если падающая волна определяется уравнениями (6); при этом U получается из известного выражения для $E_z\sec\beta$ в двумерном случае заменой k на $k\cos\beta$ и умножением на $\exp(-ikz\sin\beta)$. Действительно, условие $U=0$ на полуплоскости $y=0$, $x>0$ предполагает также, что там и $\partial U/\partial x=0$; следовательно, из (4) находим, что на экране $E_x=E_z=0$, как и требуется. Используя (11.5.24) и (11.5.26), получим в явном виде

$$U=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\sec\beta\exp\left[ik(r\cos\beta-z\sin\beta)\right]\{G(p)-G(q)\}, \tag{8}$$

где

$$p=-\sqrt{2kr\cos\beta}\cos\frac{1}{2}(\theta-\alpha),\quad q=-\sqrt{2kr\cos\beta}\cos\frac{1}{2}(\theta+\alpha). \tag{9}$$

Таким образом, согласно (4) имеем

$$\left.\begin{aligned}
E_z&=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\cos\beta\exp\left[ik(r\cos\beta-z\sin\beta)\right]\{G(p)-G(q)\},\\
H_x&=-\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp\left[ik(r\cos\beta-z\sin\beta)\right]\times\\
&\qquad\times\left\{\sin\alpha\,[G(p)+G(q)]+i\sqrt{\frac{2}{kr\cos\beta}}\sin\frac{1}{2}\alpha\cos\frac{1}{2}\theta\right\},\\
H_y&=\frac{\exp\left(-\frac{1}{4}i\pi\right)}{\sqrt{\pi}}\exp\left[ik(r\cos\beta-z\sin\beta)\right]\times\\
&\qquad\times\left\{\cos\alpha\,[G(p)-G(q)]-i\sqrt{\frac{2}{kr\cos\beta}}\sin\frac{1}{2}\alpha\sin\frac{1}{2}\theta\right\},\\
E_x&=-H_y\sin\beta,\quad E_y=H_x\sin\beta,\quad H_z=0.
\end{aligned}\right\} \tag{10}$$

При $\beta=0$ выражения (10) сразу же превращаются в соответствующие выражения (11.5.26) и (11.5.28).

Аналогичный результат можно получить таким же образом и в том случае, когда падающая волна определяется (7). Соответствующим двумерным решением при этом является решение для H-поляризации, а именно выражение (11.5.44) для H_z. Поэтому, как уже отмечалось, можно получить решение для совершенно произвольной падающей плоской волны. Кроме того, простое обобщение аргументов, приведенных в п. 11.5.2, показывает, что поле, возникающее при *любом* распределении источников, можно представить в виде спектра плоских волн. Следовательно, в принципе, решение дифракционной задачи при любом распределении источников можно найти из решений задач для отдельных плоских волн. Ниже рассматриваются два интересных случая: линейный источник, расположенный параллельно дифракционному краю (двумерная задача), и точечный источник.

§ 11.7. Дифракция волн, испускаемых локализованным источником, на полуплоскости

11.7.1. Линейный ток, параллельный дифракционному краю. Рассмотрим линейный источник, находящийся в точке T с координатами (r_0, θ_0), где $0\leqslant\theta_0\leqslant\pi$, и излучающий в вакууме E-поляризованную цилиндрическую волну вида

$$E_z^{(i)}=\sqrt{\frac{\pi}{2}}\exp\left(\frac{1}{4}i\pi\right)H_0^{(1)}(kR)\approx\frac{\exp(ikR)}{\sqrt{kR}}. \tag{1}$$

Здесь $H_0^{(1)}$— функция Ханкеля первого рода нулевого порядка, а R — расстояние, измеряемое от T (рис. 11.16).

Фактически источник представляет собой электрический ток, протекающий через T параллельно оси z и осциллирующий везде с одинаковой фазой. Хорошо известно, что (1) является основным решением двумерного волнового уравнения, определяющим расходящуюся волну, зависящую только от радиального расстояния.

Записывая (1) в виде углового спектра плоских волн, мы в сущности используем интегральное представление функции Ханкеля (см., например, [27]),

предложенное Зоммерфельдом. Так как мы требуем, чтобы каждая плоская волна спектра достигала экрана, то следует рассматривать это представление в полупространстве $y<r_0 \sin\theta_0$, т. е.

$$H^{(1)}(kR)=\frac{1}{\pi}\int\limits_{S\left(\frac{1}{2}\pi\right)} \exp\left[ikr_0\cos(\theta_0-\alpha)\right]\exp\left[-ikr\cos(\theta-\alpha)\right]d\alpha. \quad (2)$$

Здесь выбран специальный путь интегрирования $S\left(\frac{1}{2}\pi\right)$, так как (помимо того, что он справедлив для всех точек $y<r_0 \sin\theta_0$) он параллелен пути наибыстрейшего спуска, что удобно в последующем исследовании Фазовый множитель $\exp\{ikr_0(\cos\theta_0-\alpha)\}$ появляется в подынтегральном выражении (2) потому, что все плоские волны спектра должны иметь нулевую фазу в T

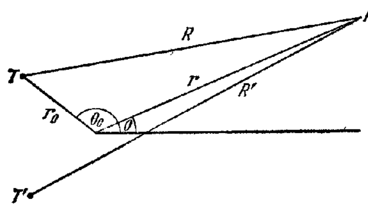


Рис. 11 16 Взаимное расположение линейного источника T и дифракционного экрана, занимающего полуплоскость.

Таким образом, очевидно, что решение дифракционной задачи для падающего поля (1) получается из решения для падающего поля $\exp\{-ikr\cos(\theta-\alpha)\}$ умножением на множитель

$$\frac{\exp\left(\frac{1}{4}i\pi\right)}{\sqrt{2\pi}}\exp\left[ikr_0\cos(\theta_0-\alpha)\right] \quad (3)$$

и интегрированием по α по пути $S\left(\frac{1}{2}\pi\right)$.

Теперь, как уже было показано в § 11.5, решение для падающего поля $\exp\{-ikr\cos(\theta-\alpha)\}$ можно записать в виде

$$E_z^{(p)}=E_z^{(pg)}+E_z^{(pd)} \quad (4)$$

(индекс p указывает на то, что падающая волна плоская), где $E_z^{(pg)}$ дается выражением (11.5.33), а (см. (11.5.19))

$$E_z^{(pd)}=\frac{i}{\pi}\int\limits_{S(0)}\frac{\sin\left(\frac{1}{2}\alpha\right)\sin\frac{1}{2}(\theta+\beta)}{\cos\alpha+\cos(\theta+\beta)}\exp(ikr\cos\beta)\,d\beta. \quad (5)$$

Преимущество такого представления $E_z^{(pd)}$ (вместо представления в виде интегралов Френеля) состоит в том, что, как мы скоро увидим, здесь отчетливо проявляется симметрия относительно α и β. Решение для падающего поля (1), таким образом, имеет вид

$$E_z=\frac{\exp\left(\frac{1}{4}i\pi\right)}{\sqrt{2\pi}}\int\limits_{S\left(\frac{1}{2}\pi\right)}E_z^{(pg)}\exp\left[ikr_0\cos(\theta_0-\alpha)\right]d\alpha+$$
$$+\frac{\exp\left(\frac{1}{4}i\pi\right)}{\sqrt{2\pi}}\int\limits_{S\left(\frac{1}{2}\pi\right)}E_z^{(pd)}\exp\left[ikr\cos(\theta_0-\alpha)\right]d\alpha. \quad (6)$$

Для того чтобы отделить здесь члены геометрической оптики от дифракционных членов, путь интегрирования по α во втором интеграле следует изменить на $S(\theta_0)$. При таком изменении пути нужно учесть полюсы величины $E_z^{(pd)}$, рассматриваемой как функция α; согласно (5) они существуют при $\cos\alpha=-\cos(\theta+\beta)$ для всех β на $S(0)$. Легко показать, что вклад вычетов,

комбинируясь с первым членом в (6), даст члены геометрической оптики, т. е.

$$E_z^{(g)} = \begin{cases} \sqrt{\frac{1}{2}\pi} \exp\left(\frac{1}{4} i\pi\right) \{H_0^{(1)}(kR) - H_0^{(1)}(kR')\} \\ \qquad\qquad \text{при } 0 \leqslant \theta \leqslant \pi - \theta_0, \\ \sqrt{\frac{1}{2}\pi} \exp\left(\frac{1}{4} i\pi\right) H_0^{(1)}(kR) \quad \text{при } \pi - \theta_0 < \theta < \pi + \theta_0, \\ 0 \qquad\qquad \text{при } \pi + \theta_0 < \theta \leqslant 2\pi, \end{cases} \tag{7}$$

где R' — расстояние, измеренное от T' — изображения T в плоскости $y = 0$ (см. рис. 11.16). Дифракционный член можно записать в виде

$$E_z^{(d)} = \frac{\exp\left(-\frac{1}{4} i\pi\right)}{\pi\sqrt{2\pi}} \int\limits_{S(0)} \int\limits_{S(0)} \frac{\sin\frac{1}{2}(\alpha+\theta_0)\sin\frac{1}{2}(\beta+\theta)}{\cos(\alpha+\theta_0)+\cos(\beta+\theta)} \times$$
$$\times \exp\left[ik(r_0\cos\alpha + r\cos\beta)\right] d\alpha\, d\beta. \tag{8}$$

Его удобно представить следующим образом:

$$E_z^{(d)} = \mathscr{Y}(\theta_0) - \mathscr{Y}(-\theta_0), \tag{9}$$

где

$$\mathscr{Y}(\theta_0) = -\frac{\exp\left(-\frac{1}{4} i\pi\right)}{4\pi\sqrt{2\pi}} \int\limits_{S(0)} \int\limits_{S(0)} \frac{\exp\left[ik(r_0\sin\alpha + r\cos\beta)\right]}{\cos\frac{1}{2}(\alpha+\beta+\theta_0+\theta)} d\alpha\, d\beta. \tag{10}$$

Окончательное преобразование должно свести $\mathscr{Y}(\theta_0)$ к однократному интегралу. Умножая числитель и знаменатель подынтегрального выражения (10) на $4\cos\frac{1}{2}(\alpha-\beta+\theta_0+\theta)$ и отбрасывая часть, являющуюся нечетной функцией β, имеем

$$\mathscr{Y}(\theta_0) = -\frac{\exp\left(-\frac{1}{4} i\pi\right)}{8\pi\sqrt{2\pi}} \int\limits_{S(0)} \int\limits_{S(0)} \left\{ \frac{1}{\cos\frac{1}{2}(\alpha+\theta_0+\theta) - \sin\left(\frac{1}{2}\beta\right)} + \right.$$
$$\left. + \frac{1}{\cos\frac{1}{2}(\alpha+\theta_0+\theta) + \sin\left(\frac{1}{2}\beta\right)} \right\} \cos\left(\frac{1}{2}\beta\right) \exp\left[ik(r_0\cos\alpha + r\cos\beta)\right] d\alpha\, d\beta. \tag{11}$$

Заменим α во втором члене подынтегрального выражения (11) на $-\alpha$ и объединим два члена. Это дает

$$\mathscr{Y}(\theta_0) = -\frac{\exp\left(-\frac{1}{4} i\pi\right)}{2\pi\sqrt{2\pi}} \int\limits_{S(0)} \int\limits_{S(0)} \frac{1}{N} \left\{ \cos\frac{1}{2}(\theta_0+\theta)\cos\left(\frac{1}{2}\alpha\right)\cos\left(\frac{1}{2}\beta\right) \times \right.$$
$$\left. \times \exp\left[ik(r_0\cos\alpha + r\cos\beta)\right] \right\} d\alpha\, d\beta, \tag{12}$$

где

$$N = (\cos\alpha - 1) + (\cos\beta - 1) -$$
$$- 4\sin\frac{1}{2}(\theta_0+\theta)\sin\left(\frac{1}{2}\alpha\right)\sin\left(\frac{1}{2}\beta\right) + 2\cos^2\frac{1}{2}(\theta_0+\theta).$$

Произведя в (12) подстановки, применяемые в методе наибыстрейшего спуска

$$\xi = \sqrt{2}\exp\left(\frac{1}{4} i\pi\right)\sin\left(\frac{1}{2}\alpha\right), \qquad \eta = \sqrt{2}\exp\left(\frac{1}{4} i\pi\right)\sin\left(\frac{1}{2}\beta\right),$$

и написав $R_1 = r_0 + r$, получим

$$\mathscr{Y}(\theta_0) = \frac{\exp\left(-\frac{1}{4}i\pi\right)}{\pi\sqrt{2\pi}} \exp(ikR_1)\cos\frac{1}{2}(\theta_0+\theta) \times$$
$$\times \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \frac{\exp\left[-k(r_0\xi^2 + r\eta^2)\right]}{\xi^2+\eta^2+2\sin\frac{1}{2}(\theta_0+\theta)\xi\eta - 2i\cos^2\frac{1}{2}(\theta_0-\theta)}\,d\xi\,d\eta. \quad (13)$$

Далее введем полярные координаты

$$\xi = \sqrt{\frac{R_1}{r_0}}\rho\cos\varphi, \qquad \eta = \sqrt{\frac{R_1}{r_0}}\rho\sin\varphi.$$

Тогда

$$\mathscr{Y}(\theta_0) = \frac{\exp\left(-\frac{1}{4}i\pi\right)}{\pi\sqrt{2\pi}} \exp(ikR_1)\cos\frac{1}{2}(\theta_0+\theta)\int\limits_0^{\infty}\rho K(\rho)\exp(-kR_1\rho^2)\,d\rho, \quad (14)$$

где

$$K(\rho) = \int\limits_0^{2\pi}\left\{\rho^2\left[\sqrt{\frac{r}{r_0}}\cos^2\varphi + \sqrt{\frac{r_0}{r}}\sin^2\varphi + 2\sin\frac{1}{2}(\theta_0+\theta)\sin\varphi\cos\varphi\right] - \right.$$
$$\left. -2i\frac{\sqrt{rr_0}}{R_1}\cos^2\frac{1}{2}(\theta_0+\theta)\right\}^{-1} d\varphi. \quad (15)$$

$K(\rho)$ можно вычислить обычным способом, полагая $z = \exp(i\varphi)$, когда путь интегрирования становится окружностью единичного радиуса и необходимо только найти вычеты окруженных полюсов. Таким образом, получим

$$K(\rho) = 2\pi\left|\sec\frac{1}{2}(\theta_0+\theta)\right|\left\{\rho^4 - 2i\rho^2 - \frac{4rr_0}{R_1^2}\cos^2\frac{1}{2}(\theta_0+\theta)\right\}^{-1/2}, \quad (16)$$

где ветвь квадратного корня лежит (для вещественных значений ρ) в четвертом квадранте комплексной плоскости. Следовательно, (10) принимает вид

$$\mathscr{Y}(\theta_0) = \pm\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\exp(ikR_1)\times$$
$$\times\int\limits_0^{\infty}\frac{\rho\exp(-kR_1\rho^2)}{\sqrt{\left\{\left[\rho^2 - \frac{i(R_1-R')}{R_1}\right]\left[\rho^2 - \frac{i(R_1+R')}{R_1}\right]\right\}}}\,d\rho. \quad (17)$$

Верхний знак берется для $\cos\frac{1}{2}(\theta_0+\theta) > 0$, нижний — для $\cos\frac{1}{2}(\theta_0+\theta) < 0$. Окончательно подстановка

$$\mu^2 = ikR_1\rho^2 + k(R_1 - R')$$

приводит к искомому результату, а именно

$$\mathscr{Y}(\theta_0) = \pm\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\exp(ikR')\int\limits_{\sqrt{k(R_1-R')}}^{\infty}\frac{\exp(i\mu^2)}{\sqrt{\mu^2+2kR'}}\,d\mu, \quad (18)$$

где верхний знак берется для $\cos\frac{1}{2}(\theta_0+\theta) > 0$, нижний — для $\cos\frac{1}{2}(\theta_0+\theta) < 0$.

Так как падающее поле (1) можно записать в виде

$$\sqrt{\frac{2}{\pi}}\exp\left(-\frac{1}{4}i\pi\right)\exp(ikR)\int\limits_{-\infty}^{\infty}\frac{\exp(i\mu^2)}{\sqrt{\mu^2+2kR}}\,d\mu, \quad (19)$$

член геометрической оптики (7), комбинируясь с дифракционным членом

(19), даст для полного поля соотношение

$$E_z = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{4} i\pi\right) \left\{ \exp(ikR) \int\limits_m^\infty \frac{\exp(i\mu^2)}{\sqrt{\mu^2 + 2kR}} d\mu - \right.$$

$$\left. - \exp(ikR') \int\limits_{m'}^\infty \frac{\exp(i\mu^2)}{\sqrt{\mu^2 + 2kR'}} d\mu \right\}, \tag{20}$$

где

$$\left.\begin{aligned} m &= -2\sqrt{\frac{krr_0}{R_1+R}} \cos\frac{1}{2}(\theta_0 - \theta) = \\ &= \mp\sqrt{k(R_1 - R)}, \quad \mp \text{ для } \cos\frac{1}{2}(\theta_0 - \theta) \lessgtr 0, \\ m' &= -2\sqrt{\frac{krr_0}{R_1+R'}} \cos\frac{1}{2}(\theta_0 + \theta) = \\ &= \mp\sqrt{k(R_1 - R')}, \quad \mp \text{ для } \cos\frac{1}{2}(\theta_0 + \theta) \lessgtr 0. \end{aligned}\right\} \tag{21}$$

Это решение в виде (20) впервые дано Макдональдом [28], получившим его с помощью преобразования предыдущего решения, предложенного Карслоу [2]. Оно очень близко к решению, найденному Зоммерфельдом для падающей плоской волны, с которым оно совпадает, если после умножения на $\sqrt{kr_0}\exp(-ikr_0)$ положить, что $r_0 \to \infty$. Решение для H-поляризации отличается от предыдущего только тем, что в уравнении (20) два члена складываются, а не вычитаются.

Если $kR_1 \gg 1$, то в неэкспоненциальном множителе подынтегрального выражения (18) μ можно заменить величиной его нижнего предела, после чего мы получим приближенный результат

$$\mathscr{Y}(\theta_0) = \pm\sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{4} i\pi\right) \frac{\exp(ikR')}{\sqrt{k(R_1 + R')}} F\left[\sqrt{k(R_1 - R')}\right]. \tag{22}$$

Таким образом, используя соответствующее приближение для $\mathscr{Y}(-\theta_0)$, можно представить поле дифракции в виде интегралов Френеля. Однако точность такого приближения недостаточна, если как источник, так и точка наблюдения находятся в пределах длины волны от дифракционного края.

Далее, если $k(R_1 - R') \gg 1$, то можно воспользоваться асимптотическим приближением (11.5.31) для (22), и мы получим

$$\mathscr{Y}(\theta_0) = \frac{\exp\left(\frac{1}{4} i\pi\right)}{2\sqrt{2\pi}} \sec\frac{1}{2}(\theta_0 + \theta) \frac{\exp(ikr_0)}{\sqrt{kr_0}} \frac{\exp(ikr)}{\sqrt{kr}}; \tag{23}$$

аналогично, если $k(R_1 - R) \gg 1$, то

$$\mathscr{Y}(-\theta_0) = \frac{\exp\left(\frac{1}{4} i\pi\right)}{2\sqrt{2\pi}} \sec\frac{1}{2}(\theta_0 - \theta) \frac{\exp(ikr_0)}{\sqrt{kr_0}} \frac{\exp(ikr)}{\sqrt{kr}}. \tag{24}$$

Поэтому для всех точек, лежащих вне двух *гипербол* $k(R_1 - R') = 1$ и $k(R_1 - R) = 1$ с осями $\theta + \theta_0 = \pi$ и $\theta - \theta_0 = \pi$ соответственно, поле дифракции совпадает с полем некоторого линейного источника, находящегося у дифракционного края. Эти гиперболы эквивалентны параболам для случая падающей цилиндрической волны, которые рассматривались в п. 11.5.3 в связи с решением для падающей плоской волны.

Наконец, следует обратить внимание на то обстоятельство, что решение (20) «*обратимо*» в том смысле, что оно не изменяется при взаимной замене переменных r_0, θ_0 и r, θ. Это, конечно, частный случай общей теоремы взаимности [19], неявно содержащейся в уравнениях Максвелла. Проведенный

анализ показывает, что в настоящем случае это связано с тем, что спектральная функция (11.5.7) симметрична относительно α и α_0.

11.7.2. **Диполь.** Простейшим точечным источником электромагнитных волн является электрический или магнитный диполь. Задачу о диполе в присутствии полуплоскости можно решить, представляя невозмущенное поле диполя как спектр (трехмерный) плоских волн и применяя результаты, найденные в § 11.6, к каждой плоской волне. Случай электрического диполя с осью, перпендикулярной дифракционному экрану, разобран таким методом Сениором [29] и кратко анализируется здесь; другие методы разобраны в [30].

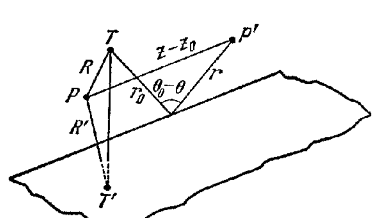


Рис. 11.17. Взаимное расположение диполя, находящегося в точке T, и дифракционного экрана, занимающего полуплоскость.

На рис. 11.17 показано (как и раньше в декартовых x, y, z и цилиндрических r, θ, z координатах) взаимное расположение диполя и дифракционного экрана, занимающего полуплоскость $y=0$, $x>0$. Диполь, находящийся в T, с координатами (x_0, y_0, z_0) или (r_0, θ_0, z_0) параллелен оси y; T' — изображение T в плоскости $y=0$, а R и R' — расстояние точки наблюдения P от T и T' соответственно *).

При подходящем выборе величины дипольного момента невозмущенное поле диполя (см. § 2.2) можно представить в виде

$$\mathbf{E}=\left(\frac{\partial^2\Pi}{\partial x\,\partial y}, \frac{\partial^2\Pi}{\partial y^2}+k^2\Pi, \frac{\partial^2\Pi}{\partial y\,\partial z}\right),\quad \mathbf{H}=ik\left(\frac{\partial\Pi}{\partial z},\ 0,\ -\frac{\partial\Pi}{\partial x}\right), \tag{25}$$

где

$$\Pi=\frac{\exp(ikR)}{kR}. \tag{26}$$

Требуемое решение (25) в виде плоских волн получается из формулы **)

$$\frac{\exp(ikR)}{kR}=\frac{i}{2\pi}\int\limits_{S\left(\frac{1}{2}\pi\right)}\int\limits_{S\left(\frac{1}{2}\pi\right)}\cos\beta\exp\left[-ik\left\{(x-x_0)\cos\alpha\cos\beta+\right.\right.$$
$$\left.\left.+(y-y_0)\sin\alpha\cos\beta-(z-z_0)\sin\beta\right\}\right]d\alpha\,d\beta. \tag{27}$$

Теперь поступают следующим образом: сначала записывают поле, возникшее в присутствии дифракционного экрана в результате падения на него отдельных плоских волн, определяемых подынтегральным выражением в (27), а затем выполняют интегрирование.

Учитывая последующее интегрирование, удобно использовать (см. (5)) основные решения для падающих плоских волн в виде (11.5.8) (для E-поляризации) и соответствующий вид решений для H-поляризации. Эти выражения видоизменяют так, чтобы способом, изложенным в § 11.6, получить трехмерные решения, соответствующие падению плоских волн типа (27). Тогда компонента вектора $\mathbf{E}$, параллельная диполю, имеет вид

$$E_y=E_y^{(g)}+E_y^{(d)}; \tag{28}$$

здесь $E_y^{(g)}$ — поле геометрической оптики, а

$$E_y^{(d)}=\frac{ik}{2\pi}\{A(\theta_0)+A(-\theta_0)\}-\frac{i}{2\pi k}\left(\frac{\partial^2}{\partial y\,\partial y_0}-k^2\right)B(\theta_0)+$$
$$+\frac{i}{2\pi k}\left(\frac{\partial^2}{\partial y\,\partial y_0}+k^2\right)B(-\theta_0), \tag{29}$$

*) P' — проекция точки P на плоскость, проходящую через T и T' перпендикулярно границе дифракционного экрана. (*Прим. ред.*)

**) Это непосредственно следует из формулы, приведенной в [31].

где

$$A(\theta_0)=\frac{ik}{2\pi}\int\limits_{S(0)}\int\limits_{S\left(\frac{1}{2}\pi\right)}\int\limits_{S(0)}\cos\beta\cos\frac{1}{2}(\alpha+\gamma+\theta-\theta_0)\times$$

$$\times\exp\left[ik\{r\cos\gamma\cos\beta+r_0\cos\alpha\cos\beta+(z-z_0)\sin\beta\}\right]d\alpha\, d\beta\, d\gamma, \quad (30)$$

$$B(\theta_0)=\frac{ik}{2\pi}\int\limits_{S(0)}\int\limits_{S\left(\frac{1}{2}\pi\right)}\int\limits_{S(0)}\cos\beta\sec\frac{1}{2}(\alpha+\gamma+\theta-\theta_0)\times$$

$$\times\exp\left[ik\{r\cos\gamma\cos\beta+r_0\cos\alpha\cos\beta+(z-z_0)\sin\beta\}\right]d\alpha\, d\beta\, d\gamma. \quad (31)$$

С помощью анализа, аналогично данному Карслоу [2] и Макдональдом [28], можно показать, что

$$A(\theta_0)=\frac{\pi}{\sqrt{rr_0}}\cos\frac{1}{2}(\theta-\theta_0)H_0^{(1)}(kR_1), \quad (32)$$

где

$$R_1^2=(r+r_0)^2+(z-z_0)^2; \quad (33)$$

$$B(\theta_0)=\pm\pi k\int\limits_{\pm m}^{\infty}H_1^{(1)}(kR\,\mathrm{ch}\,\mu)\,d\mu, \quad \pm \text{ для } \theta-\theta_0\lessgtr\pi, \quad (34)$$

$$B(-\theta_0)=\pm\pi k\int\limits_{\pm m'}^{\infty}H_1^{(1)}(kR'\,\mathrm{ch}\,\mu)\,d\mu, \quad \pm \text{ для } \theta+\theta_0\lessgtr\pi, \quad (35)$$

$$m=\mathrm{Arsh}\left\{2\frac{\sqrt{rr_0}}{R}\cos\frac{1}{2}(\theta-\theta_0)\right\}, \quad m'=\mathrm{Arsh}\left\{2\frac{\sqrt{rr_0}}{R'}\cos\frac{1}{2}(\theta+\theta_0)\right\}. \quad (36)$$

Эти результаты вместе с формулой

$$\frac{1}{2}i\int\limits_{-\infty}^{\infty}H_1^{(1)}(kR\,\mathrm{ch}\,\mu)\,d\mu=\frac{\exp(ikR)}{kR} \quad (37)$$

позволяют переписать (28) в виде

$$E_y=\frac{ik}{\sqrt{rr_0}}\cos\left(\frac{1}{2}\theta\right)\cos\left(\frac{1}{2}\theta_0\right)H_0^{(1)}(kR_1)-\frac{1}{2}i\left(\frac{\partial^2}{\partial y\,\partial y_0}-k^2\right)\mathscr{Y}+$$

$$+\frac{1}{2}i\left(\frac{\partial^2}{\partial y\,\partial y_0}+k^2\right)\mathscr{Y}', \quad (38)$$

где

$$\mathscr{Y}=\int\limits_{-m}^{\infty}H_1^{(1)}(kR\,\mathrm{ch}\,\mu)\,d\mu, \qquad \mathscr{Y}'=\int\limits_{-m'}^{\infty}H_1^{(1)}(kR'\,\mathrm{ch}\,\mu)\,d\mu. \quad (39)$$

Остальные компоненты поля точно выражаются через $\mathscr{Y}$ и $\mathscr{Y}$ следующим образом:

$$E_x=-\frac{ik}{\sqrt{rr_0}}\sin\left(\frac{1}{2}\theta\right)\cos\left(\frac{1}{2}\theta_0\right)H_0^{(1)}(kR_1)-\frac{1}{2}i\frac{\partial^2\mathscr{Y}}{\partial x\,\partial y_0}+\frac{1}{2}i\frac{\partial^2\mathscr{Y}'}{\partial x\,\partial y_0}, \quad (40)$$

$$E_z=-\frac{1}{2}i\frac{\partial^2\mathscr{Y}}{\partial z\,\partial y_0}+\frac{1}{2}i\frac{\partial^2\mathscr{Y}'}{\partial z\,\partial y_0}, \quad (41)$$

$$H_x=\frac{k(z-z_0)}{R_1\sqrt{rr_0}}\cos\left(\frac{1}{2}\theta\right)\cos\left(\frac{1}{2}\theta_0\right)H_1^{(1)}(kR_1)+\frac{1}{2}k\frac{\partial\mathscr{Y}}{\partial z_0}+\frac{1}{2}k\frac{\partial\mathscr{Y}'}{\partial z_0}, \quad (42)$$

$$H_y=\frac{k(z-z_0)}{R_1\sqrt{rr_0}}\sin\left(\frac{1}{2}\theta\right)\cos\left(\frac{1}{2}\theta_0\right)H_1^{(1)}(kR_1), \quad (43)$$

$$H_z=-\frac{1}{2}k\frac{\partial\mathscr{Y}}{\partial x_0}-\frac{1}{2}k\frac{\partial\mathscr{Y}'}{\partial x_0}. \quad (44)$$

Интересно отметить, что полное поле имеет не равную нулю компоненту вектора **H**, параллельную диполю, так что результат анализа нельзя выразить через один вектор Герца.

Здесь опять можно (при обычных ограничениях) представить решение в виде интегралов Френеля. Если $kR_1 \gg 1$, то нетрудно получить следующее асимптотическое приближение:

$$\mathscr{Y} = -\frac{2\exp\left(\frac{1}{4}i\pi\right)}{k}\sqrt{\frac{2}{\pi R_1(R+R_1)}}\times$$
$$\times\exp(ikR)F\left\{-2\sqrt{\frac{krr_0}{R+R_1}}\cos\frac{1}{2}(\theta-\theta_0)\right\}. \quad (45)$$

Аналогичный результат можно получить и для $\mathscr{Y}'$.

§ 11.8. Другие задачи

В настоящем параграфе коротко рассматривается несколько других дифракционных задач.

11.8.1. Две параллельные полуплоскости. Задача о дифракции света на краях двух параллельных полуплоскостей, перпендикулярных к плоскости, проходящей через их края, разбирается с помощью метода, изложенного в настоящей главе для случая одиночной полуплоскости *).

Рассмотрим E-поляризацию для точно такого же случая, как и в § 11.5.1, но теперь дифракционное препятствие состоит из *двух* экранов. Пусть один из них (экран 1) занимает полуплоскость $y=0$, $x>0$, другой (экран 2) — полуплоскость $y=-a$, $x>0$. Удобно ввести дополнительные координаты r' и θ', измеряемые от края $(0, -a)$ экрана 2.

Дифрагированные поля, вызванные индуцированными токами в экранах 1 и 2, можно записать соответственно в виде

$$E_z^{(s1)} = \int_C P_1(\cos\alpha)\exp[ikr\cos(\theta\mp\alpha)]\,d\alpha, \quad \mp \text{ для } y \gtrless 0, \quad (1)$$

$$E_z^{(s2)} = \int_C P_2(\cos\alpha)\exp[ikr'\cos(\theta'\mp\alpha)]\,d\alpha, \quad \mp \text{ для } y \gtrless -a. \quad (2)$$

Непрерывность $H_x^{(s1)}$ и $H_x^{(s2)}$ при переходе через области $y=0$, $x<0$ и $y=-a$, $x<0$ обеспечивается, если предположить, что $P_1(\mu)$ и $P_2(\mu)$ свободны от сингулярностей в области ниже пути интегрирования, показанного на рис. 11.7. Кроме того, граничное условие, требующее исчезновения E_z на обоих экранах, приводит к интегральным уравнениям

$$\int_{-\infty}^{\infty}\frac{P_1(\mu)}{\sqrt{1-\mu^2}}\exp(ikx\mu)\,d\mu + \int_{-\infty}^{\infty}\frac{P_2(\mu)}{\sqrt{1-\mu^2}}\exp\left(ika\sqrt{1-\mu^2}\right)\exp(ikx\mu)\,d\mu =$$
$$= -\exp(-ikx\mu_0), \quad (3)$$

$$\int_{-\infty}^{\infty}\frac{P_1(\mu)}{\sqrt{1-\mu^2}}\exp\left(ika\sqrt{1-\mu^2}\right)\exp(ikx\mu)\,d\mu + \int_{-\infty}^{\infty}\frac{P_2(\mu)}{\sqrt{1-\mu^2}}\exp(ikx\mu)\,d\mu =$$
$$= -\exp\left(ika\sqrt{1-\mu_0^2}\right)\exp(-ikx\mu_0), \quad (4)$$

которые должны быть справедливыми для $x>0$.

Положив

$$P_1(\mu)+P_2(\mu)=Q_1(\mu), \qquad P_1(\mu)-P_2(\mu)=Q_2(\mu), \quad (5)$$

*) Решения для падающей плоской волны сначала были получены методом, предложенным Швингером (он упоминается в § 11.1), Хейнсом [10] и Вайнштейном [32].

получим, складывая и вычитая (3) и (4),

$$\int_{-\infty}^{\infty} \frac{Q_1(\mu)}{\sqrt{1-\mu^2}} \{1+\exp(ika\sqrt{1-\mu^2})\} \exp(ikx\mu)\, d\mu =$$
$$= -\{1+\exp(ika\sqrt{1-\mu_0^2})\} \exp(-ikx\mu_0), \tag{6}$$

$$\int_{-\infty}^{\infty} \frac{Q_2(\mu)}{\sqrt{1-\mu^2}} \{1-\exp(ika\sqrt{1-\mu^2})\} \exp(ikx\mu)\, d\mu =$$
$$= -\{1-\exp(ika\sqrt{1-\mu_0^2})\} \exp(-ikx\mu_0) \tag{7}$$

для $x>0$.

Уравнения (6) и (7) сходны с уравнением (11.5.2), и для получения решений, аналогичных (11.5.4), пути интегрирования следует замкнуть бесконечной полуокружностью выше вещественной оси. С этой целью выбирается та ветвь $\sqrt{1-\mu^2}$, у которой мнимая часть положительна. Тогда требуемое решение уравнения (6) примет вид

$$\frac{Q_1(\mu)}{\sqrt{1-\mu^2}} \{1+\exp(ika\sqrt{1-\mu^2})\} = \frac{i}{2\pi} \{1+\exp(ika\sqrt{1-\mu_0^2})\} \frac{U(\mu)}{U(-\mu_0)} \frac{1}{\mu+\mu_0}, \tag{8}$$

где $U(\mu)$ — произвольная функция μ, свободная от сингулярностей в полуплоскости выше пути интегрирования и стремящаяся к нулю при $|\mu|\to\infty$.

Остается показать, как следует выбрать $Q_1(\mu)$ и $U(\mu)$, чтобы они удовлетворяли (8). Напомним, что у $Q_1(\mu)$ нет сингулярностей ниже выбранного пути интегрирования. Задача сводится к тому, чтобы представить коэффициент при $Q_1(\mu)$ в соотношении (8) в виде

$$\frac{1+\exp(ika\sqrt{1-\mu^2})}{\sqrt{1-\mu^2}} = U_1(\mu)\, L_1(\mu), \tag{9}$$

где $U_1(\mu)$ не имеет сингулярностей и нулей в полуплоскости выше пути интегрирования и по абсолютной величине возрастает там до бесконечности; $L_1(\mu)$ должна обладать такими же характеристиками в полуплоскости ниже пути интегрирования. О возможности подобного представления известно из общей теории Винера и Хопфа [7]. Хейнс [10] получил выражения для $U_1(\mu)$ и $L_1(\mu)$ в явном виде. В таком случае имеем

$$Q_1(\mu) = \frac{i}{2\pi} \sqrt{1-\mu_0^2}\, U_1(\mu_0) \frac{1}{L_1(\mu)(\mu+\mu_0)}, \tag{10}$$

где использовано соотношение $U_1(\mu)=L_1(-\mu)$, входящее в неявном виде в (9) (с точностью до произвольного постоянного множителя).

Аналогично, если

$$\frac{1-\exp(ika\sqrt{1-\mu^2})}{\sqrt{1-\mu^2}} = U_2(\mu)\, L_2(\mu), \tag{11}$$

то

$$Q_2(\mu) = \frac{i}{2\pi} \sqrt{1-\mu_0^2}\, U_2(\mu_0) \frac{1}{L_2(\mu)(\mu+\mu_0)}. \tag{12}$$

Полное дифрагированное поле находится сложением (1) и (2); так, например, при $y>0$ имеем

$$E_z^{(s)} = \int_C \{P_1(\cos\alpha) + P_2(\cos\alpha)\exp(ika\sin\alpha)\} \exp[ikr\cos(\theta-\alpha)]\, d\alpha =$$
$$= \frac{1}{2}\int_C \{Q_1(\cos\alpha)(1+\exp(ika\sin\alpha)) + Q_2(\cos\alpha)(1-\exp(ika\sin\alpha))\} \times$$
$$\times \exp[ikr\cos(\theta-\alpha)]\, d\alpha = \int_C P(\cos\alpha)\exp[ikr\cos(\theta-\alpha)]\, d\alpha, \tag{13}$$

где

$$P(\mu)=\tfrac{1}{2}\sqrt{1-\mu^2}\,\{Q_1(\mu)U_1(\mu)L_1(\mu)+Q_2(\mu)U_2(\mu)L_2(\mu)\}=$$

$$=\frac{i}{4\pi}\frac{\sqrt{1-\mu_0^2}\sqrt{1-\mu^2}}{\mu+\mu_0}\{U_1(\mu_0)U_1(\mu)+U_2(\mu_0)U_2(\mu)\}. \qquad (14)$$

Здесь снова следует отметить симметрию $P(\mu)$ относительно μ и μ_0. Кроме того, когда $a=0$, имеем $U_1(\mu)=\sqrt{\frac{2}{1+\mu}}$, $U_2(\mu)=0$, и (14) сводится к (11.5.6).

11.8.2. Бесконечный набор параллельных полуплоскостей, расположенных ступеньками. В этой задаче [9] рассматривается бесконечный ряд дифракционных экранов, причем n-й экран занимает полуплоскость $y=na$, $x>nb$, где $n=0, \pm1, \pm2, \ldots$

Как и раньше, для падающей E-поляризованной плоской волны дифрагированное поле, обусловленное индуцированным током в m-м экране, имеет вид

$$E_z^{(sm)}=\int_C P_m(\cos\alpha)\exp[-ikm(a\cos\alpha\pm b\sin\alpha)]\exp[ikr\cos(\theta\mp\alpha)]\,d\alpha, \qquad (15)$$

где верхний знак берется для $y>ma$, а нижний — для $y<ma$. Все функции $P_m(\mu)$ должны быть свободны от сингулярностей ниже пути интегрирования в плоскости $\mu=\cos\alpha$.

Из граничных условий на n-м экране следует интегральное уравнение

$$\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{P_m(\mu)}{\sqrt{1-\mu^2}}\exp(ika|n-m|\sqrt{1-\mu^2})\exp[ik(x-mb)\mu]\,d\mu=$$

$$=-\exp(-ikx\mu_0)\exp\left(-ikna\sqrt{1-\mu_0^2}\right), \qquad (16)$$

которое должно выполняться для $x>nb$.

Из периодичности задачи ясно, что

$$P_m(\mu)=P_0(\mu)\exp\left[-ikm\{b\mu_0+a\sqrt{1-\mu_0^2}\}\right]; \qquad (17)$$

полагая $n-m=q$, приведем (16) к виду

$$\sum_{q=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{P_0(\mu)}{\sqrt{1-\mu^2}}\exp\left[ikq\{b\mu_0+a\sqrt{1-\mu_0^2}\}\right]\exp(ika|q|\sqrt{1-\mu^2})\times$$

$$\times\exp(ik(qb+x)\mu)\,d\mu=-\exp(-ikx\mu_0) \qquad (18)$$

для $x>0$.

Бесконечную сумму по q можно представить в замкнутой форме, если снова написать интегральное уравнение, решаемое при помощи теоремы вычетов Коши. Как и в предыдущей задаче, необходимо представить некоторую функцию в виде двух множителей. Один из них должен быть свободен от сингулярностей и нулей в верхней полуплоскости и по абсолютной величине возрастать там до бесконечности, а другой должен обладать такими же характеристиками в нижней полуплоскости. Более подробно с этими множителями можно познакомиться в статьях Карлсона и Хейнса, на которые мы уже ссылались.

11.8.3. Полоса. Другая интересная задача, вводящая в заблуждение своей кажущейся простотой, относится к дифракции на бесконечно длинной, идеально проводящей плоской полосе с параллельными краями или к дифракции на «дополнительном экране» в виде щели в бесконечной плоскости. Было предложено несколько способов решения этой задачи [6, 16, 33—36], но ни один из них не давал решения в замкнутом виде. Ниже показано, как в случае нормального падения плоской волны метод дуального интегрального уравнения [37, 38] использовался для получения в решении первых двух членов разложения в степенной ряд по ka, где $2a$ — ширина полосы.

Для полосы, занимающей область $y=0$, $|x|<a$, при нормальном падении H-поляризованной плоской волны интегральные уравнения (11.4.19) и (11.4.20) принимают вид

$$\int_{-\infty}^{\infty} P(\mu)\exp(ikx\mu)\,d\mu = 1 \quad \text{для} \quad |x|<a, \tag{19}$$

$$\int_{-\infty}^{\infty} \frac{P(\mu)}{\sqrt{1-\mu^2}}\exp(ikx\mu)\,d\mu = 0 \quad \text{для} \quad |x|>a, \tag{20}$$

или, поскольку из симметрии задачи следует, что $P(\mu) = P(-\mu)$,

$$\int_{0}^{\infty} P(\mu)\cos(kx\mu)\,d\mu = \frac{1}{2} \quad \text{для} \quad |x|<a, \tag{21}$$

$$\int_{0}^{\infty} \frac{P(\mu)}{\sqrt{1-\mu^2}}\cos(kx\mu)\,d\mu = 0 \quad \text{для} \quad |x|>a. \tag{22}$$

Будем искать решение в виде

$$\frac{P(\mu)}{\sqrt{1-\mu^2}} = \sum_{m=0}^{\infty} c_m \frac{J_{2m+1}(ka\mu)}{\mu}, \tag{23}$$

так как (22) удовлетворяется каждым членом этого ряда. Подстановка в (21) показывает, что c_m должно быть таким, чтобы

$$\sum_{m=0}^{\infty} c_m \Phi_m = \frac{1}{2} \quad \text{для} \quad |x|<a, \tag{24}$$

где

$$\Phi_m = \int_{0}^{\infty} \frac{\sqrt{1-\mu^2}}{\mu} J_{2m+1}(ka\mu)\cos(kx\mu)\,d\mu. \tag{25}$$

Теперь можно показать, что для $|ka| \ll 1$

$$\Phi_m = i\int_{0}^{\infty} J_{2m+1}(ka\mu)\cos(kx\mu)\,d\mu + O(ka) = \frac{i\cos\left[(2m+1)\arcsin\frac{x}{a}\right]}{ka\sqrt{1-x^2/a^2}} + O(ka). \tag{26}$$

Отсюда в первом приближении получим вместо (24)

$$\frac{i}{\sqrt{1-x^2/a^2}} \sum_{m=0}^{\infty} c_m \cos\left[(2m+1)\arcsin\frac{x}{a}\right] = \frac{1}{2}ka, \tag{27}$$

что дает

$$c_0 = \frac{ka}{2i}, \qquad c_m = 0 \quad \text{для} \quad m = 1,\ 2,\ 3,\ \ldots \tag{28}$$

Таким образом, в этом приближении

$$\frac{P(\mu)}{\sqrt{1-\mu^2}} = \frac{ka}{2i}\frac{J_1(ka\mu)}{\mu} \tag{29}$$

и плотность тока, выведенная из (11.4.16), равна

$$J_x(x) = \frac{cka}{2\pi i}\int_{0}^{\infty} \frac{J_1(ka\mu)}{\mu}\cos(kx\mu)\,d\mu = \frac{cka}{2\pi i}\sqrt{1-x^2/a^2}. \tag{30}$$

Следующее приближение, учитывающее члены $(ka)^3$, значительно сложнее. Различные авторы независимо получили для него одинаковые выражения ([35, 38], см. также [13]).

11.8.4. Другие задачи. Существует еще целый ряд интересных задач, близких к рассмотренным нами, и они также допускают решение. Однако мы вынуждены только упомянуть здесь о них.

Задача о дифракции на двумерном клине, который превращается в полуплоскость, если внешний угол клина равен 2π, решена много лет тому назад [3, 39]. В это решение входит угловой спектр периода $2\pi/n$, где π/n — внешний угол клина.

Задача о дифракции на полуплоскости, находящейся между двумя различными однородными средами, была впервые рассмотрена Хансоном [40]. Задача решается методом, изложенным в настоящей главе; она используется в теории распространения радиоволн над поверхностью Земли [41].

Были выполнены два исследования дифракции на полуплоскости в менее идеализированных условиях. В первом [42] была введена конечная, хотя и большая проводимость, что вынуждало использовать приближенные граничные условия; во втором [43] предполагалось, что плоскость является идеально проводящей, но имеет конечную, хотя и малую толщину.

§ 11.9. Единственность решения

Мы видели в § 11.2, что общую задачу о дифракции, которой мы касались, можно сформулировать следующим образом: дано поле $\mathbf{E}^{(i)}$, падающее на идеально проводящую поверхность S, требуется найти поле $\mathbf{E}^{(s)}$, обусловленное электрическим током, распределенным по S, которое имеет такую величину, что и его тангенциальная составляющая на S равняется тангенциальной составляющей $\mathbf{E}^{(i)}$, взятой со знаком минус.

Конечно, существенно необходимо, чтобы этой формулировке соответствовала единственность решения*). Однако показать, что действительно не существует другого поля $\mathbf{E}^{(s)}$, удовлетворяющего поставленным условиям, совсем не просто, особенно когда допускаются бесконечные размеры S и учитываются поля, содержащие плоские волны. По-видимому, только сравнительно недавно появились удовлетворительные доказательства единственности [46] (см. также [45, 47]), хотя она давно молчаливо принималась.

Дополнительные трудности возникают в особом, хотя и наиболее общем случае дифракционной задачи, в которой можно предположить, что дифракционное препятствие имеет бесконечно острый край. (Фактически такой задачей и ограничивалось содержание настоящей главы.) Причина возникновения дополнительных сложностей заключается в том, что это решение содержит, как мы видели, сингулярности на краю и тем самым нарушается предположение, необходимое для только что упомянутого доказательства единственности решения.

Легко видеть, что если допускаются произвольные сингулярности по краю, то можно получить бесконечную последовательность решений путем дифференцирования [48]. Например, дифференцирование по x решения для E-поляризации в задаче с полуплоскостью (см. (11.5.22)) дает по существу новое выражение, также удовлетворяющее волновому уравнению и обращающееся в нуль на экране. Дифференцирование (11.5.22) по y дает выражение, которое, очевидно, служит решением для H-поляризации, но отличается от (11.5.44).

Каждое дифференцирование вносит сингулярность более высокого порядка в решение для дифракционного края. Очевидно, для обеспечения единственности решения следует точно определить некоторые ограничения, налагаемые на характер сингулярности. Целесообразные ограничения и различные способы, которыми их можно сформулировать, явились предметом целого ряда статей [49]. Со всеми деталями, относящимися к этому вопросу, читатель

*) Установить *существование* решения, вероятно, менее важно, так как такое заключение делается, когда решение найдено в любом частном случае (см. также [44, 45]).

может познакомиться непосредственно по оригинальным статьям. Однако, вообще говоря, решения, содержащие сингулярность наинизшего порядка, следует считать ответом на физическую задачу, и это значит, что в действительности исключаются любые сингулярности более высоких порядков, чем $r^{-1/2}$ при $r \to 0$ на дифракционном крае. В частности, можно утверждать, что решение будет единственным, если имеет место интегрируемость индуцированного тока по дифракционной поверхности и исчезновение у края соответствующей компоненты, нормальной к краю. Поведение компонент векторов **E** и **H** вблизи края можно вывести из этих условий.

Наконец, убедившись в возможности получить бесконечное число «решений», можно задать вопрос, почему метод, использованный в настоящей главе, приводит к единственному решению, которое к тому же, если применять упомянутый выше критерий, оказывается правильным. Это объясняется допущением, что компоненты поля в плоскости экрана можно выразить в виде сходящихся интегралов Фурье, вследствие чего они не имеют также сингулярностей слишком высокого порядка.

ЛИТЕРАТУРА

1. Sommerfeld, Math. Ann. **47**, 317 (1896).
2. H. S. Carslaw, Proc. Lond. Math. Soc. **30**, 121 (1899).
3. H. M. Macdonald, Electric Waves, Cambr. Univ. Press, 1902.
4. T. J. I'A. Bromwich, Proc. Lond. Math. Soc. **14**, 450 (1916).
5. C. J. Bouwkamp, Dissertation, Groningen, 1941; J. Meixner, W. Andrejewski, Ann. d. Physik 7, 157 (1950).
6. Rayleigh, Phil. Mag. **43**, 259 (1897).
7. E. C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Clarendon Press, Oxford, 1937. (Е. Титчмарш, Введение в теорию интегралов Фурье, Гостехиздат, 1948.)
8. E. T. Copson, Quart. J. Maths **17**, 19 (1946).
9. J. F. Carslon, A. E. Heins. Quart. Appl. Maths **4**, 313 (1947); **5**, 82 (1947).
10. A. E. Heins, Quart. Appl. Maths **6**, 157, 215 (1948).
11. H. Levine, J. Schwinger, Phys. Rev. **73**, 383 (1948).
12. J. W. Miles, J. Appl. Phys. **20**, 760 (1949).
13. C. J. Bouwkamp, в сб. «Rep. Progr. Phys.», Phys. Soc. London, 1954, Vol. 17, pp. 35, 73.
14. H. Levine, J. Schwinger, Phys. Rev. **74**, 958 (1948); **75**, 1423 (1949).
15. G. Wolfsohn, в кн. «Handbuch der Physik», Springer, Berlin, 1928, Vol. 20, p. 263, 273.
16. B. B. Baker, E. T. Copson, The Mathematical Theory of Huygens' Principle, 2nd ed. Clarendon Press, Oxford, 1950.
17. H. Höhl, A. W. Maue, K. Westpfahl, в кн. «Handbuch der Physik», Springer, Berlin, 1961, Vol. 25/1.
18. H. G. Booker, J. IEE **93**, Pt. IIIA, 620 (1946).
19. L. G. H. Huxley, The Principles and Practice of Waveguides, Cambr. Univ. Press, 1947, p. 284, § 7.17.
20. E. T. Whittaker, G. N. Watson, A Course of Modern Analysis, Cambr. Univ. Press, 1927, p. 397. (Э. Уиттекер, Г. Ватсон, Курс современного анализа, Гостехиздат, 1933.)
21. R. A. Rankin, Phil. Trans. Roy. Soc. **A241**, 457 (1949).
22. G. G. Stokes, Trans. Cambr. Phil. Soc. **10**, 105 (1864).
23. W. Braunbek, G. Laukien, Optik **9**, 174 (1952).
24. J. Savornin, Ann. de Physique **11**, 129 (1939).
25. C. W. Horton, R. B. Watson, J. Appl. Phys. **21**, 16 (1950).
26. B. N. Harden. Proc. IEE **99**, Pt. III, 229 (1952); R. D. Kodis, J. Appl. Phys. **23**, 249 (1952); R. V. Row, J. Appl. Phys. **24**, 1448 (1953).
27. J. A. Stratton, Electromagnetic Theory, McGraw-Hill, N. Y., 1941. (Дж. Стрэттон, Теория электромагнетизма, Гостехиздат, 1948.)
28. H. M. Macdonald, Proc. Lond. Math. Soc. **14**, 410 (1915).
29. T. B. A. Senior, Quart. J. Mech. Appl. Maths **6**, 101 (1953).
30. A. E. Heins, Trans. IRE, **AP-4**, 294 (1956); B. D. Woods, Quart. J. Mech. Appl. Maths. **10**, 90 (1957); W. E. Williams, Quart. J. Mech. Appl. Maths **10**, 210 (1957).
31. H. Weyl, Ann. d. Physik **60**, 481 (1919).
32. Л. А. Вайнштейн, Известия АН СССР, Сер. физ., **12**, 144, 166 (1948).
33. K. Schwarzchild, Math. Ann. **55**, 177 (1902); P. M. Morse, P. J. Rubenstein, Phys. Rev. **54**, 895 (1938); A. Sommerfeld, Optics, Acad. Press, N. Y., 1954. (А. Зоммерфельд, Оптика, ИЛ, 1953.)

34. S. Skavlem, Arch. Math. Naturvid. 51, 61 (1951), E. B. Moullin, F. M. Phillips, Proc. IEE 99, Pt. IV, 137 (1952).
35. R. Müller, K. Westpfahl, Z. Phys. 134, 245 (1953).
36. P. C. Clemmow, Trans, IRE, AP-4, 282 (1956), S. N. Karp, A. Russek. J. Appl. Phys. 27, 886 (1956,); R. F. Millar, Proc. Cambr Phil. Soc. 54, 479, 497 (1958).
37. E. Groschwitz, H. Hönl, Z. f. Phys. 131, 305 (1952); H. Hönl, E. Zimmer, Z. f. Phys. 135, 196 (1953).
38. C. J. Tranter, Quart. J. Mech. Appl. Maths 7, 317 (1954).
39. H. S. Carslaw, Proc. Lond. Math. Soc. 18, 291 (1919).
40. E. T. Hanson, Phil. Trans. Roy. Soc. A237, 35 (1938).
41. P. C. Clemmow, Phil. Trans. Roy. Soc. A246, 1 (1953).
42. T. B. A. Senior, Proc. Roy. Soc. A213, 436 (1952).
43. D. S. Jones, Proc. Roy. Soc. A217, 153 (1953).
44 C. Müller, Math. Ann. 123, 345 (1951).
45. W. K. Saunders, Proc. Nat. Acad. Sci. USA 38, 342 (1952).
46. F. Rellich, Jahr, Deut. Math. Ver. 53, 57 (1943).
47. A. Sommerfeld, Partial Differential Equations in Physics, Acad. Press, N. Y., 1949, § 29. (А. Зоммерфельд, Дифференциальные уравнения в частных производных физики, ИЛ, 1950.)
48. C. J. Bouwkamp, Physica 12, 467 (1946).
49. J. Meixner, Ann. d. Physik 6, 1 (1949); A. W. Maue, Z. f. Phys. 126, 601 (1949); E. T. Copson, Proc. Roy. Soc. A202, 277 (1950); D. S. Jones, Quart. J. Mech. Appl. Maths 3, 420 (1950); A. E. Heins, S. Silver, Proc. Cambr. Phil. Soc. 51, 149 (1955); 54, 131 (1958).

ГЛАВА 12

ДИФРАКЦИЯ СВЕТА НА УЛЬТРАЗВУКОВЫХ ВОЛНАХ

В гл. 1 и 2 было показано, что распространение электромагнитных волн можно изучать либо с помощью уравнений Максвелла, дополненных материальными уравнениями, либо с помощью некоторых интегральных уравнений, учитывающих поляризационные свойства среды. В частности, каждый из этих методов можно применять также для изучения распространения света в среде, плотность которой зависит от пространственных координат и от времени. Первый метод широко применялся в прошлом, тогда как последний только сравнительно недавно стал использоваться в таких исследованиях. В настоящей главе мы воспользуемся методом интегральных уравнений для решения задач, связанных с дифракцией света в прозрачной однородной среде, возмущенной проходящими ультразвуковыми волнами. Однако сначала полезно дать качественное описание этого явления и кратко изложить соответствующие теоретические работы, основанные на применении дифференциальных уравнений Максвелла.

§ 12.1 Качественное описание явления и краткое изложение теорий, основанных на применении дифференциальных уравнений Максвелла

12.1.1. Качественное описание явления. Ультразвуковые волны — это звуковые волны, частоты которых выше частот, воспринимаемых ухом человека. Угловая частота ультразвуковых волн, генерируемых в лабораториях, находится приблизительно между 10^5 *сек*$^{-1}$ и $3\cdot 10^9$ *сек*$^{-1}$. Первая величина представляет предельную частоту, воспринимаемую нашим ухом. Соответствующий диапазон для волн Λ, очевидно, зависит от скорости v этих волн в среде, в которой они распространяются. Например, в воде $v=1{,}2\cdot 10^5$ *см/сек* и указанные выше частоты соответствуют диапазону длин волн *) от $\Lambda=7{,}5$ *см* до $\Lambda=2{,}5\cdot 10^{-4}$ *см*.

В 1921 г. Бриллюэн [2] предсказал, что при освещении жидкости, в которой распространяются упругие волны небольшой длины, видимым светом возникнет дифракция, подобная дифракции на решетке. Для того чтобы убедиться в этом, рассмотрим жидкость, находящуюся между двумя бесконечными плоскостями $y=0$ и $y=d$, и допустим, что плоская упругая волна длиной Λ движется в ней вдоль положительного направления оси x. Это создает периодическую слоистость в веществе вдоль оси x, причем расстояние между последовательными плоскостями максимальной плотности равно Λ.

Рассмотрим падающую монохроматическую плоскую световую волну с угловой частотой ω и длиной волны в веществе $\bar{\lambda}$, причем волновая нормаль лежит в плоскости xy и составляет угол $\bar{\theta}$ с осью y (рис. 12.1). Далее пусть $\bar{\varphi}$ обозначает угол, который дифрагировавший луч образует с осью y. Так как скорость v упругих волн всегда значительно меньше скорости света, можно в первом приближении принять, что слоистая структура вещества, вызванная

*) О методах генерации ультразвуковых волн и их многочисленных применениях см., например, [1].

ультразвуковой волной, неподвижна. В таком случае направления $\bar{\varphi}$, в которых имеется заметная интенсивность, определяются условием, требующим, чтобы оптическая разность хода между лучами от двух соседних плоскостей, находящихся на расстоянии Λ, равнялась целому кратному $\bar{\lambda}$. Это условие определяет зависимость между $\bar{\lambda}$, $\bar{\theta}$ и направлениями распространения $\bar{\varphi}_l$ волн разных порядков в дифракционном спектре, т. е.

$$BC-AD=\Lambda(\sin\bar{\varphi}_l-\sin\bar{\theta})=l\bar{\lambda}, \qquad i=0,\ \pm1,\ \pm2,\ \ldots \tag{1}$$

Здесь AB и CD — части волновых фронтов, связанных с преломленными и дифрагировавшими лучами. Удобно выразить (1) через углы θ, φ и длину волны λ вне среды. Если воспользоваться в (1) законом преломления

$$\frac{\sin\bar{\theta}}{\sin\theta}=\frac{\sin\bar{\varphi}}{\sin\varphi}=\frac{\bar{\lambda}}{\lambda},$$

то получим

$$\Lambda(\sin\varphi_l-\sin\theta)=l\lambda,$$
$$l=0,\ \pm1,\ \pm2,\ \ldots \tag{2}$$

Рис. 12.1. Ультразвуковые волны как дифракционная решетка.

Из (2) находим для углового расстояния между соседними порядками

$$\sin\varphi_l-\sin\varphi_{l-1}\approx\varphi_l-\varphi_{l-1}=\lambda/\Lambda.$$

Таким образом, при данной длине волны λ угловое расстояние уменьшается с увеличением Λ. Если Λ достаточно велико, то главные линии располагаются так тесно, что инструмент, с помощью которого ведется наблюдение, не сможет их разрешить. Этим обстоятельством объясняется то, что дифракционные явления не наблюдались, пока видимым светом освещались обычные звуковые волны.

Почти через десять лет после предсказаний Бриллюэна Дебай и Сирс [3] и независимо от них Люка и Бикар [4] обнаружили дифракцию света на ультразвуковых волнах. С тех пор многие исследователи изучали это явление в различных экспериментальных условиях при изменении одного или нескольких из следующих параметров: (а) угла падения света θ, (б) длины Λ ультразвуковой волны, (в) длины волны λ падающего света, (г) амплитуды ультразвуковых волн, (д) ширины d ультразвукового пучка.

Естественно, что положение спектров различных порядков на экране, их число и относительная интенсивность зависят от одного или нескольких из перечисленных выше факторов *). На приводимом ниже рис. 12.3 показано для типичного случая число порядков, появляющихся с каждой стороны прошедшего пучка света, при разных углах падения θ на ультразвуковую решетку. Обычное экспериментальное устройство, применяемое для изучения дифракционных спектров, схематически изображено на рис. 12.2.

На этой стадии знакомства с дифракцией света на ультразвуковых волнах удобно ввести и определить некоторые принятые в настоящей главе символы и знаки.

Концентрация молекул (атомов) в среде обозначается через $N(\mathbf{r},t)$. В изотропной однородной среде, в которой плоская упругая волна распространяется

---

*) Полуколичественная теория зависимости дифракционных спектров от этих факторов изложена в [5].

в положительном направлении оси x, $N(\mathbf{r}, t)$ можно записать в виде *)

$$N(\mathbf{r}, t) = N_0 [1 + \Delta \cos (K_x - \Omega t)], \quad (3)$$

где N_0 — средняя концентрация молекул в среде, $N_0\Delta$ (обычно порядка $10^{-4}N_0$) — амплитуда упругой волны (в единицах N_0), $K = 2\pi/\Lambda$ — волновое число (величина волнового вектора) упругой волны и $\Omega = Kv$ — угловая частота ультразвукового возмущения. В такой среде диэлектрическая

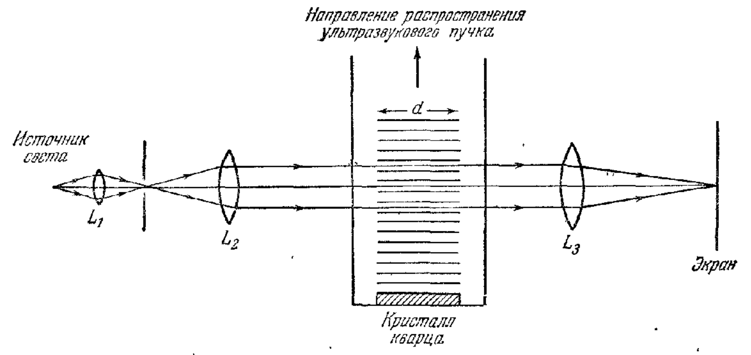


Рис. 12.2. Схема экспериментальной установки для наблюдения дифракции света на ультразвуковых волнах.

проницаемость ε также зависит от пространственных координат и от времени. Эту зависимость можно представить в виде

$$\varepsilon = \varepsilon_0 + \varepsilon_1 \cos (Kx - \Omega t). \quad (4)$$

Конечно, между ε_1 и Δ существует связь, которую можно записать следующим образом:

$$\varepsilon_1 = \gamma\Delta. \quad (5)$$

Если мы примем закон Лорентц — Лоренца (см. (2.3.17))

$$\frac{\varepsilon - 1}{\varepsilon + 2} \frac{1}{N_i} = \text{const}, \quad (6)$$

то, дифференцируя логарифм (6) и вспоминая, что Δ и $\varepsilon_1/\varepsilon_0$ значительно меньше единицы, получим

$$\varepsilon_1 \left[\frac{1}{\varepsilon_0 - 1} - \frac{1}{\varepsilon_0 + 2}\right] - \frac{N_0\Delta}{N_0} = 0$$

или

$$\gamma = \frac{1}{3}(\varepsilon_0 - 1)(\varepsilon_0 + 2), \quad (7)$$

т. е. для большинства жидкостей величина γ порядка единицы.

Мы считаем также, что

$$n = \sqrt{\varepsilon_0}, \quad k = \frac{2\pi}{\lambda}, \quad \bar{k} = nk, \quad \delta = \frac{\Delta\Lambda^2}{\lambda^2}, \quad \xi = \frac{\Lambda}{\lambda}\sin\theta, \quad \beta = \frac{\pi\lambda}{n\Lambda^2}. \quad (8)$$

Наконец, углы θ, φ и т. д. измеряются по часовой стрелке от положительного направления оси y к направлению, по которому распространяется свет (см. рис. 12.1). Предполагается также, что $0 \leqslant \theta < \pi/2$.

*) Для простоты мы рассмотрим только плоские бегущие ультразвуковые волны. Дифракция света на стоячих волнах также изучалась экспериментально (см. [1]); соответствующее обобщение теории не встречает затруднений.

Следует отметить, что в эксперименте очень трудно получить совершенно плоские ультразвуковые волны. Однако волновые фронты можно считать плоскими в областях, линейные размеры которых значительно больше Λ.

12.1.2. Краткое изложение теорий, основанных на уравнениях Максвелла. В областях, свободных от токов и зарядов, уравнения Максвелла для немагнитной непроводящей среды с диэлектрической проницаемостью ε, зависящей от пространственных координат и времени, имеют вид

$$\operatorname{rot}\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{H}}{\partial t}, \qquad \operatorname{rot}\boldsymbol{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \tag{9а}$$

$$\operatorname{div}\mathbf{H} = 0, \qquad \operatorname{div}\mathbf{D} = 0. \tag{9б}$$

Исключая $\mathbf{H}$ из (9а) и используя соотношения $\mathbf{D} = \varepsilon\mathbf{E}$, $\operatorname{div}\mathbf{D} = 0$ и тождество $\operatorname{rot}\operatorname{rot} \equiv -\nabla^2 + \operatorname{grad}\operatorname{div}$, найдем, что

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\varepsilon\mathbf{E}) = \nabla^2\mathbf{E} + \operatorname{grad}(\mathbf{E}\cdot\operatorname{grad}\ln\varepsilon). \tag{10}$$

Если теперь воспользоваться выражением (4) и рассматривать $\mathbf{E}$ как суперпозицию плоских волн с длинами $\bar{\lambda} \approx \lambda/n$, то мы найдем, что второй член в правой части (10) приблизительно в $\varepsilon_1(\lambda/\Lambda)$ раз больше первого. Так как в обычных условиях эксперимента ε_1 и λ/Λ значительно меньше единицы, то можно пренебречь этим членом в (10), и мы получим

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}(\varepsilon\mathbf{E}) = \nabla^2\mathbf{E}. \tag{11}$$

Рассмотрим теперь с физической точки зрения ситуацию, изображенную на рис. 12.1. Допустим, что падающая монохроматическая плоская электромагнитная волна линейно поляризована и ее электрический вектор перпендикулярен плоскости падения (E-поляризация), т. е. направлен вдоль оси z. В этом случае из предыдущего следует, что внутри среды компоненты E_x E_y вектора $\mathbf{E}$ будут малыми величинами порядка $\varepsilon_1(\lambda/\Lambda)$ E_z, и поэтому ими можно пренебречь. Следовательно, вне среды E_x и E_y также пренебрежимо малы.

Из симметрии задачи ясно, что E_z не зависит от координаты z, и поэтому, воспользовавшись (4) и (11), получим для E_z уравнение

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left\{\left[\varepsilon_0 + \frac{1}{2}\varepsilon_1(\exp\left[i(Kx - \Omega t)\right] + \exp\left[-i(Kx - \Omega t)\right]\right]E_z\right\} = 0. \tag{12}$$

Чтобы решить его, допустим, что E_z имеет вид

$$E_z = \sum_l V_l(y)\exp\{i\,[(k\sin\theta + lK)\,x - (\omega + l\Omega)\,t]\}, \tag{13}$$

где суммирование ведется по всем целым значениям l (положительным, отрицательным и нулевому).

Подставляя (13) в (12) и приравнивая коэффициенты при каждой экспоненте нулю, получим следующие рекуррентные соотношения для $V_l(y)$:

$$V_l''(y) + \{\varepsilon_0 c^{-2}(\omega + l\Omega)^2 - (k\sin\theta + lK)^2\}\,V_l(y) =$$
$$= -\frac{1}{2}\varepsilon_1 c^{-2}(\omega + l\Omega)^2\,\{V_{l-1}(y) + V_{l+1}(y)\} \qquad (l = 0,\ \pm 1,\ \pm 2,\ \ldots), \tag{14}$$

где штрих у $V_l(y)$ обозначает дифференцирование по y. Эти уравнения следует решать при граничных условиях *)

$$\left.\begin{aligned} &V_0(0) = B, \text{ амплитуда падающей световой волны,} \\ \text{и}\quad &V_l(0) = 0 \text{ для всех } l \neq 0. \end{aligned}\right\} \tag{15}$$

Не решая (14), можно видеть, что (13) представляет суперпозицию волн с частотами $\omega_l = \omega + l\Omega$ ($l = 0,\ \pm 1,\ \pm 2,\ \ldots$). Кроме того, x-компонента вол-

*) Эти граничные условия правильны только при ничтожно малой интенсивности отраженных волн. В данной задаче это так и есть, потому что угол падения θ, для которого амплитуды дифрагировавших волн достаточно велики, не превышает 3° (см. также п. 12.2.4).

нового вектора волны с частотой ω_l равна $k\sin\theta + lK$. Следовательно, синус угла φ_l, который нормаль к волне с частотой ω_l образует с осью y вне рассеивающей среды, равняется

$$\sin\varphi_l = \frac{c(k\sin\theta + lK)}{\omega + l\Omega} \approx \sin\theta + l\frac{\lambda}{\Lambda}, \quad \text{так как} \quad \frac{\Omega}{\omega} \ll 1, \tag{16}$$

что находится в согласии с (2). Далее интенсивность в спектре определенного порядка l можно принять равной $|V_l(d)|^2$.

Решим (14), предположив сначала, что $V_0 \gg V_{\pm1} \gg V_{\pm2} \ldots$ Напомним, что нужно рассматривать только те решения (14), которые соответствуют световым волнам, движущимся в направлении увеличения y, поэтому, принимая, что в (14) все V_l, кроме V_0, равны нулю и используя (15), получим в первом приближении

$$V_0^{(0)}(y) = B\exp\left[i\sqrt{(\varepsilon_0 - \sin^2\theta)}\,ky\right] + O(\delta^2). \tag{17}$$

Аналогично, полагая в (14) все V_l, кроме $V_{\pm1}$ и V_0, равными нулю, находим после несложных расчетов

$$V_{\pm1}(y) = \frac{1}{4}\gamma\delta\,\frac{1 - \exp\left[-2i\beta\left(\xi \pm \frac{1}{2}\right)y\right]}{\xi \pm \frac{1}{2}}\,V_0^{(0)}(y). \tag{18}$$

Здесь использованы также соотношения (5) и (8). Если теперь подставить (18) в уравнение (14) для $l = 0$ и для $l = \pm 2$, то мы получим поправочный член к $V_0^{(0)}(y)$ и выражение для $V_{\pm2}(y)$, которые в данном приближении пропорциональны δ^2. Таким же способом получаются выражения для интенсивности в спектре любого порядка l в виде разложения в ряд по возрастающим степеням δ. Решение в виде такого степенного ряда впервые было получено Бриллюэном [6], применившим, правда, несколько более сложный анализ, чем приведенный здесь. Дэвид [7], следуя изложенному выше методу, получил в явном виде выражения для интенсивностей линий в спектрах первого и второго порядка, предполагая, что интенсивности в спектрах высших порядков пренебрежимо малы. Эти формулы упоминаются ниже (см. (12.2.38)).

Приближение Бриллюэна (а также Дэвида) удобно при $\delta \ll 1$ или $\delta/\xi \ll 1$, когда степенной ряд сходится быстро. Если эти условия выполняются, то очевидно, что заметная интенсивность будет наблюдаться только в спектрах нескольких первых порядков.

Объяснение одновременного появления спектров многих порядков и приближенные выражения, определяющие их интенсивности, были впервые получены Раманом и Натом *) [8]. Они решили уравнение (14) следующим образом. Полагая в (14)

$$V_l(y) = \exp\left(\frac{1}{2}il\pi\right)\exp\left[i\bar{k}(\cos\bar{\theta})\,y\right]U_l(y) \tag{19}$$

и вспоминая, что $\Omega/\omega \approx 10^{-5}$ (или меньше этой величины), $\bar{k} = nk$ и $k\sin\theta = \bar{k}\sin\bar{\theta}$, получим следующие рекуррентные соотношения для $U_l(y)$:

$$U_l''(y) + 2i\bar{k}\cos\bar{\theta}\,U_l'(y) - (2l\bar{k}K\sin\bar{\theta} + l^2K^2)\,U_l(y) + \\ + \frac{1}{2}i(\bar{k})^2\varepsilon_1\varepsilon_0^{-1}\left(U_{l+1}(y) - U_{l-1}(y)\right) = 0 \qquad (l = 0, \pm1, \pm2, \ldots). \tag{20}$$

Вводя новую переменную

$$\chi = \frac{1}{2}y\bar{k}\varepsilon_1\varepsilon_0^{-1}\sec\bar{\theta},$$

*) Отметим, что Раман и Нат, следуя в двух своих первых статьях работе Рэлея [9], относящейся к фазовым решеткам, нашли, что амплитуды дифрагированных волн можно представить в виде функции Бесселя. В этом методе рассматриваются изменения только фазы плоской световой волны при ее пересечении ультразвукового пучка.

приведем последнее уравнение к виду

$$2U_l'(\chi)+U_{l+1}(\chi)-U_{l-1}(\chi)=$$
$$=\left(\frac{1}{2}i\varepsilon_0^{-1}\varepsilon_1\right)\sec^2\bar{\theta}U_l''(\chi)-2i\varepsilon_0\varepsilon_1^{-1}\{2lK(\bar{k})^{-1}\sin\bar{\theta}+l^2K^2(\bar{k})^{-2}\}U_l(\chi). \quad (21)$$

Здесь штрих у U означает дифференцирование по χ. Так как $\lambda/d \ll 1$, первый член в правой части обычно порядка $\varepsilon_1 U$, т. е. $10^{-4}\ U$, и его можно отбросить. Кроме того, если мы, следуя Раману и Нату, допустим, что и первый член в фигурных скобках равен нулю, то оставшийся ряд уравнений представляет собой рекуррентные соотношения (см., например, [10]), которым удовлетворяют бесселевы функции целого порядка. Используя граничные условия (15), получим для интенсивности в спектре порядка l выражение $B^2J_l^2(^1/_2\varepsilon_1\varepsilon_0^{-1}kd\sec\bar{\theta})$.

Следует отметить, что приближение, сделанное Раманом и Натом, основано главным образом на пренебрежении l^2/δ и $l\xi/\delta$ для всех l. Поэтому, если δ достаточно велико по сравнению с единицей, то это приближение будет хорошо описывать интенсивности в низших порядках. Однако выражения, содержащие функции Бесселя, дают завышенную интенсивность в высших порядках. Это было показано численными расчетами интенсивностей для трех значений параметра δ, выполненными Экстерманом и Ванье [11]. В работе этих авторов решение уравнения (12) в конце концов определяется уравнениями, в основном подобными соотношениям (12.2.18) и (12.2.19).

Наконец, заметим, что были получены решения (21) в виде степенных рядов по возрастающим степеням $1/\delta$ [12, 13]. Эти ряды, по-видимому, имеют довольно ограниченное применение, так как сходятся очень медленно. Другая трактовка, основанная на уравнениях Максвелла, в которой дифракция рассматривается как граничная задача, дана Вагнером [14].

§ 12.2. Рассмотрение дифракции света на ультразвуковых волнах методом интегральных уравнений

В § 2.4 было отмечено, что интегральные уравнения (2.4.4) для эффективного электрического поля $\mathbf{E}'(\mathbf{r}, t)$ и соответствующая формула (2.4.5) для $\mathbf{H}'$ эквивалентны уравнениям Максвелла для изотропных немагнитных веществ. Это справедливо, если допустить, что плотность среды не зависит от времени, однако полученный результат легко распространить и на более общий случай, когда такая зависимость от времени существует. Как и раньше, мы будем считать среду немагнитной и непроводящей.

Напомним, что сущность метода интегральных уравнений заключается в том, что влияние среды на распространение электромагнитной волны считается эквивалентным действию электрических диполей, находящихся в вакууме, причем дипольный момент, индуцированный в каком-нибудь физически бесконечно малом элементе объема $d\mathbf{r}'$ с линейными размерами, значительно меньшими *) λ, пропорционален полю $\mathbf{E}'(\mathbf{r}', t)$, действующему на этот объем, и числу заключенных в нем молекул (атомов). Связанный с таким диполем в $\mathbf{r}'$ вектор Герца

$$\Pi_e=\alpha N\left(\mathbf{r}',\ t-\frac{R}{c}\right)\frac{\mathbf{E}'(t-R/c,\ \mathbf{r}')}{R}d\mathbf{r}'$$

позволяет получить поле в точке $\mathbf{r}$ в момент t с помощью операции (см. (2.2.43))

$$-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\operatorname{grad}\operatorname{div}.$$

Здесь различные символы имеют то же значение, что и в § 2.4; следовательно,

*) Такие элементы объема, значительно большие объема отдельного атома (молекулы) но с линейными размерами, малыми по сравнению с λ, почти всегда можно выбрать для оптических длин волн.

$R=|\mathbf{r}-\mathbf{r}'|$ и оператор grad div действует на переменные $\mathbf{r}(x, y, z)$. В таком случае, аргументируя так же, как и при выводе уравнения (2.4.4), можно получить следующее интегральное уравнение для $\mathbf{E}$ в среде *):

$$\mathbf{E}'(\mathbf{r}t)=\mathbf{E}^{(i)}(\mathbf{r}t)+$$
$$+\alpha\iiint\left\{-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\operatorname{grad}\operatorname{div}\right\}\left\{N(\mathbf{r}', t-R/c)\frac{\mathbf{E}'(\mathbf{r}', t-R/c)}{R}\right\}d\mathbf{r}'. \quad (1)$$

Как и в (2.4.4), интегрирование производится по всей среде, за исключением небольшой области, занятой атомом в точке наблюдения $\mathbf{r}$ (x, y, z).

Это основное интегральное уравнение рассматриваемой здесь теории. Когда оно решается относительно $\mathbf{E}'$ во всех точках внутри среды, поле вне среды рассчитывается путем добавления к падающему полю $\mathbf{E}^{(i)}$ $(\mathbf{r}, t)$ поля диполя $\mathbf{E}^{(d)}$ $(\mathbf{r}, t)$, определяемого интегралом в (1), но взятым по всей области, занятой средой. Следует отметить, что в противоположность обычному методу, требующему составления уравнений Максвелла для среды и вакуума, такое рассмотрение распространения света в среде позволяет избежать явного введения граничных условий на преломляющих поверхностях. Вместо этого в данном методе в пределы интегрирования вводятся размеры среды. Кроме того, изменения плотности среды учитываются в уравнениях Максвелла косвенным путем через диэлектрическую проницаемость ε, тогда как в интегральные уравнения (1) функция плотности $N(\mathbf{r}, t)$ входит явно.

Уравнение (1) справедливо только при определенных ограничениях. Во-первых, поляризуемость α, рассчитанная на одну молекулу, вообще говоря, зависит от частоты поля $\mathbf{E}'$, так что оно должно быть строго монохроматичным. Однако если мы не находимся слишком близко к резонансным частотам, то изменения α с частотой внешнего поля малы. Следовательно, при условии, что все частотные компоненты $\mathbf{E}'$ близки друг другу, мы по-прежнему можем использовать уравнение (1). Даже если падающее поле $\mathbf{E}^{(i)}$ строго монохроматично, поле $\mathbf{E}'$, действующее на молекулу и вызывающее образование диполей, не обязательно монохроматично, если $N(\mathbf{r}, t)$ зависит от времени, от термического возбуждения или от других причин, вызывающих разупорядочение. Ширина полосы частот зависит от вариации $N(\mathbf{r}, t)$ во времени. Таким образом, уравнением (1) можно пользоваться с уверенностью только тогда, когда изменение во времени N происходит медленно по сравнению с изменением **) $\mathbf{E}^{(i)}$. К счастью, это ограничение не является очень серьезным, так как в задачах о рассеянии и дифракции света требуемое условие почти всегда выполняется.

Далее мы считали α скаляром; такое предположение полностью оправдано для атомов и молекул, обладающих особой симметрией. Оно справедливо также и в более общем случае, когда молекулы беспорядочно ориентированы, что уже отмечалось в § 2.3. Наконец, здесь предполагается, что поглощение света средой ничтожно мало; его можно учесть, допустив, что α комплексно.

Теперь воспользуемся методом интегрального уравнения для изучения дифракции света на ультразвуковых волнах в жидкости [15, 16].

12.2.1. Интегральное уравнение для случая *E*-поляризации. Вернемся опять к физической ситуации, рассмотренной в п. 12.1.1, и допустим, что падающий свет линейно поляризован, причем его электрический вектор перпендикулярен к плоскости падения; тогда компоненты электрического вектора $\mathbf{E}^{(i)}(\mathbf{r}, t)$ падающей световой волны выражаются соотношениями (вещественная

---

*) Уравнение (1) отличается от соотношения (2.4.4) тем, что оператор rot rot заменен в последнем на $\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\operatorname{grad}\operatorname{div}\right)$ и что плотность N, которая теперь зависит от пространственных координат и времени, следует брать, подобно $\mathbf{E}'$, в момент времени $t-R/c$.

**) Использование соотношения $\mathbf{D}=\varepsilon\mathbf{E}$ в уравнениях Максвелла для неоднородной среды также оправдано только тогда, когда на N налагаются аналогичные ограничения.

часть, как обычно, представляет собой физическую величину)

$$\left.\begin{aligned} &E_x^{(i)}=E_y^{(i)}=0, \\ &E_z^{(i)}=B\exp\left[i\left(kx\sin\theta+ky\cos\theta-\omega t\right)\right]. \end{aligned}\right\} \tag{2}$$

Сказанное выше позволяет сделать вывод, что эффективное поле $\mathbf{E}'(\mathbf{r}, t)$ в среде почти параллельно оси z, и, значит, можно принять $E'_x=E'_y=0$ и векторное интегральное уравнение (1) для $\mathbf{E}'$ сведется к единственному интегральному уравнению для E'_z. Напомним, что $N(\mathbf{r}, t)$ определяется выражением (12.1.3), и поэтому интегральное уравнение для E'_z примет вид

$$E'_z=(\mathbf{r},\ t)=B\exp\left[i\left(kx\sin\theta+ky\cos\theta-\omega t\right)\right]+ \\ +\frac{\tau_0}{4\pi}\iiint\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}\right)\left[\frac{1}{R}\left\{1+\frac{1}{2}\Delta\left(\exp\left\{i\left[Kx'-\Omega\left(t-\frac{R}{c}\right)\right]\right\}+\right.\right.\right. \\ \left.\left.\left.+\exp\left\{-i\left[Kx'-\Omega\left(t-\frac{R}{c}\right)\right]\right\}\right)\right\}E'_z\left(\mathbf{r}',\ t-\frac{R}{c}\right)\right]d\mathbf{r}', \tag{3}$$

где для удобства введено обозначение *)

$$4\pi N_0\alpha=\tau_0.$$

Из (2.3.17) находим, что

$$\tau_0=\frac{3\left(n^2-1\right)}{n^2+2}. \tag{4}$$

12.2.2. Пробное решение интегрального уравнения. Так как все плоскости, перпендикулярные к оси z, физически эквивалентны, возьмем в качестве пробного решения нашего интегрального уравнения (3) выражение вида **)

$$E'_z=\sum_{l,\,m}N_{lm}\exp\left[-i\left(\omega_{lm}t-p_lx-q_my\right)\right], \tag{5}$$

где l и m — целые числа (положительные, отрицательные и нуль). Уравнение (5), очевидно, описывает двойную бесконечную совокупность плоских волн. Такая форма возможного решения следует из существования многократных отражений и преломлений в бесконечной пластинке слоистого вещества с плоскими параллельными поверхностями. Вскоре мы увидим, что различные неизвестные N_{lm}, ω_{lm}, p_l и q_m определяются из условия, что (5) удовлетворяет интегральному уравнению (3).

Чтобы решить (3), необходимо вычислить интегралы, определяющие $\mathcal{Y}(\omega,\ p,\ q)$, а именно

$$\mathcal{Y}(\omega,\ p,\ q)\exp\left\{-i\left(\omega t-px-qy\right)\right\}= \\ =\frac{1}{4\pi}\iiint\limits_{V'}\left[-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}\right]\frac{\exp\left\{-i\left[\omega\left(t-\frac{R}{c}\right)-px'-qy'\right]\right\}}{R}\,dx'\,dy'\,dz'= \\ =\frac{1}{4\pi}\iiint\limits_{V'}\exp\left[-i\left(\omega t-px'-qy'\right)\right]\left\{\frac{\omega^2}{c^2}+\frac{\partial^2}{\partial z^2}\right\}\frac{\exp\left(\frac{i\omega R}{c}\right)}{R}\,dx'\,dy'dz'. \tag{6}$$

Здесь $\omega^2>c^2p^2$. Если точка наблюдения x, y, z находится вне рассеивающей среды, то объем V' занимает всю среду ($-\infty<x'<\infty$, $0\leqslant y'\leqslant d$, $-\infty<z'<<\infty$). Если же точка наблюдения находится внутри рассеивающей среды, интеграл берется по тому же объему, за исключением небольшой сферы радиуса a (который в конечном итоге стремится к нулю) вокруг точки наблюдения.

*) Макроскопическая величина $\tau(\mathbf{r}, t)=4\pi\alpha N(\mathbf{r}, t)$ иногда называется *коэффициентом рассеяния* среды.

**) Здесь не должно возникать недоразумения, связанного с употреблением сходных символов для амплитуд N_{lm} и концентраций молекул $N(\mathbf{r}, t)$, так как, начиная с этого места, последняя величина больше не будет входить в явном виде в наши уравнения.

Полагая

$$x_1 = x' - x, \qquad y_1 = y' - y, \qquad z_1 = z' - z,$$

можно переписать (6) (после сокращения на множитель $\exp(-i\omega t)$) в виде

$$\mathscr{Y}(\omega, p, q) = \mathscr{Y}_1 + \mathscr{Y}_2, \tag{7}$$

где

$$\mathscr{Y}_1 = \frac{1}{4\pi} \iiint_{V_1} \left\{ \exp\left[i(px_1 + qy_1)\right] \frac{\partial^2}{\partial z_1^2} \left(\frac{\exp \frac{i\omega R}{c}}{R} \right) \right\} dx_1\, dy_1\, dz_1, \tag{8}$$

$$\mathscr{Y}_2 = \frac{1}{4\pi} \frac{\omega^2}{c^2} \iiint_{V_1} \left\{ \exp\left[i(px_1 + qy_1)\right] \left(\frac{\exp \frac{i\omega R}{c}}{R} \right) \right\} dx_1\, dy_1 dz_1, \tag{9}$$

причем теперь рассеивающая среда занимает объем V_1, т. е. $-\infty < x_1 < \infty$, $-y \leqslant y_1 \leqslant d - y$, $-\infty < z_1 < \infty$. Эти интегралы вычисляются в приложении 9; они равны следующим величинам.

а) Если (x, y, z) находится в рассеивающей среде *), то

$$\mathscr{Y}(\omega, p, q) = \\ = \frac{1}{\sigma(\omega, p, q)} - \frac{\omega^2}{2c^2} \frac{\exp\{-ig(\omega, p, q)\, y\}}{g(\omega, p, q)\,[\omega^2 c^{-2} - p^2]^{1/2}} + \frac{\omega^2}{2c^2} \frac{\exp\{-ih(\omega, p, q)(y - d)\}}{h(\omega, p, q)\,[\omega^2 c^{-2} - p^2]^{1/2}}; \tag{10а}$$

б) если (x, y, z) находится за рассеивающей средой, то

$$\mathscr{Y}(\omega, p, q) = \frac{\omega^2 \exp[-ig(\omega, p, q)\, y]}{2c^2 g(\omega, p, q)\,[\omega^2 c^{-2} - p^2]^{1/2}} (\exp[ig(\omega, p, q)\, d] - 1); \tag{10б}$$

в) если (x, y, z) находится перед рассеивающей средой (т. е. с той же стороны, откуда падает свет), то

$$\mathscr{Y}(\omega, p, q) = \frac{\omega^2 \exp[-ih(\omega, p, q)\, y]}{2c^2 h(\omega, p, q)\,[\omega^2 c^{-2} - p^2]^{1/2}} (\exp[ih(\omega, p, q)\, d] - 1). \tag{10в}$$

В этих выражениях

$$\left. \begin{aligned} &\sigma(\omega, p, q) = 3(p^2 + q^2 - \omega^2 c^{-2})(p^2 + q^2 + 2\omega^2 c^{-2})^{-1}, \\ &g(\omega, p, q) = q - [\omega^2 c^{-2} - p^2]^{1/2}, \quad h(\omega, p, q) = q + [\omega^2 c^{-2} - p^2]^{1/2}. \end{aligned} \right\} \tag{11}$$

Если теперь подставить (5) в (3) и использовать соотношения (6) и (10а), то простым способом получим

$$- \sum_{l,m} N_{lm} \exp[(-i)(\omega_{lm} t - p_l x - q_m y)] + \\ + B \exp[(-i)(\omega t - kx \sin\theta - ky \cos\theta)] + \\ + \tau_0 \sum_{l,m} N_{lm} \left[\frac{\exp[(-i)(\omega_{lm} t - p_l x - q_m y)]}{\sigma(\omega_{lm}, p_l, q_m)} + \right. \\ + \frac{1}{2} \Delta \sum_{+,-} \frac{\exp[(-i)\{(\omega_{lm} \pm \Omega)\, t - (p_l \pm K)\, x - q_m y\}]}{\sigma(\omega_{lm} \pm \Omega, p_l \pm K, q_m)} - \\ - \frac{\omega_{lm}^2 \exp[(-i)\{\omega_{lm} t - p_l x - [\omega_{lm}^2 c^{-2} - p_l^2]^{1/2} y\}]}{2c^2 g(\omega_{lm}, p_l, q_m)\,[\omega_{lm}^2 c^{-2} - p_l^2]^{1/2}} - \\ - \frac{1}{2} \Delta \sum_{+,-} \frac{(\omega_{lm} \pm \Omega)^2 \exp\left[(-i)\left\{(\omega_{lm} \pm \Omega)\, t - (p_l \pm K)\, x - [c^{-2}(\omega_{lm} \pm \Omega)^2 - (p_l \pm K)^2]^{1/2} y\right\}\right]}{2c^2 g(\omega_{lm} \pm \Omega, p_l \pm K, q_m)\,[c^{-2}(\omega_{lm} \pm \Omega)^2 - (p_l \pm K)^2]^{1/2}} + \\ + \frac{\omega_{lm}^2 \exp[(-i)\{\omega_{lm} t - p_l x + [\omega_{lm}^2 c^{-2} - p_l^2]^{1/2} y\}] \exp[ih(\omega_{lm}, p_l, q_m)\, d]}{2c^2 h(\omega_l, p_l, q_m)\,[\omega_{lm}^2 c^{-2} - p_l^2]^{1/2}} + \\ + \frac{1}{2} \Delta \sum_{+,-} \frac{(\omega_{lm} \pm \Omega)^2 \exp[(-i)\{(\omega_{lm} \pm \Omega)\, t - (p_l \pm K)\, x + [c^{-2}(\omega_{lm} \pm \Omega)^2 - (p_l \pm K)^2]^{1/2} y\}]}{2c^2 h(\omega_{lm} \pm \Omega, p_l \pm K, q_m)} \times \\ \left. \times \frac{\exp[ih(\omega_{lm} \pm \Omega, p_l \pm K, q_m)\, d]}{[c^{-2}(\omega_{lm} \pm \Omega)^2 - (p_l \pm K)^2]^{1/2}} \right] = 0, \tag{12}$$

*) В § 12.2 во всех выражениях берется положительный квадратный корень, если специально не оговорен другой знак.

где знак суммирования $\sum\limits_{+,-}$ перед любым выражением должен интерпретироваться следующим образом:

$$\sum_{+,-} F(a \pm b,\ c \pm d) \equiv F(a+b,\ c+d)+F(a-b,\ c-d).$$

Для того чтобы (12) удовлетворялось в любой момент времени и в любой точке внутри рассеивающей среды, коэффициент при каждой экспоненте, отличающейся от всех остальных по любой из переменных $(x,\ y,\ t)$, должен независимо от остальных равняться нулю. Из (12) видно, что изменения ω_{lm} происходят с шагом Ω и всегда сопровождаются изменением p_l на K. Однако коэффициенты при y в различных экспонентах остаются либо неизменными (q_m), либо представляют собой одну и ту же функцию соответствующих ω_{lm} и p_l. Следовательно, можно принять, что ω_{lm} зависит только от индекса l. Кроме того, поскольку можно допустить без потери общности, что ω_0 равна частоте ω падающего света, то

$$\left.\begin{aligned} &\omega_0=\omega, && \omega_l=\omega+l\Omega, \\ &p_0=k\sin\theta, && p_l=k\sin\theta+lK \end{aligned}\right\} \quad (l=0,\ \pm1,\ \pm2,\ \ldots). \qquad \begin{aligned}(13\text{а})\\(13\text{б})\end{aligned}$$

Используя эти соотношения в (12) и перегруппировывая различные члены, получим

$$\sum_{l,m}\left[N_{lm}\left(\frac{\tau_0}{\sigma_{lm}}-1\right)+\right.$$
$$\left.+\frac{1}{2}\Delta\left(\frac{\tau_0}{\sigma_{lm}}\right)(N_{l-1,m}+N_{l+1,m})\right]\exp\left[(-i)(\omega_l t-p_l x-q_m y)\right]+$$
$$+\sum_l (B\delta_{l,0}-G_l)\exp\left[(-i)(\omega_l t-p_l x-[\omega_l^2c^{-2}-p_l^2]^{1/2}y)\right]+$$
$$+\sum_l H_l\exp\left[(-i)(\omega_l t-p_l x+[\omega_l^2c^{-2}-p_l^2]^{1/2}y)\right]\equiv 0, \qquad (14)$$

где $\delta_{l,l'}$ — символ Кронекера *) (т. е. $\delta_{l,l'}=0$, если $l\neq l'$ и $\delta_{l,l}=1$), а G_l и H_l определяются следующим образом:

$$G_l=\tau_0\omega_l^2\sum_m\left[N_{lm}+\frac{1}{2}\Delta(N_{l-1,m}+N_{l+1,m})\right]\{2c^2g_{lm}[\omega_l^2c^{-2}-p_l^2]^{1/2}\}^{-1}, \qquad (15)$$

$$H_l=\tau_0\omega_l^2\sum_m\left[N_{lm}+\frac{1}{2}\Delta(N_{l-1,m}+N_{l+1,m})\right]\times$$
$$\times\left[\exp(ih_{lm}d)\right]\{2c^2h_{lm}[\omega_l^2c^{-2}-p_l^2]^{1/2}\}^{-1}. \qquad (16)$$

Здесь использованы также сокращения (см. (11))

$$\sigma_{lm}=\sigma(\omega_l,\ p_l,\ q_m),\quad g_{lm}=g(\omega_l,\ p_l,\ q_m)\quad\text{и}\quad h_{lm}=h(\omega_l,\ p_l,\ q_m). \qquad (17)$$

Приравнивая нулю коэффициенты при каждой экспоненте в (14), можно получить следующую систему уравнений для допустимых значений q_m и амплитуд N_{lm}:

$$N_{lm}\left(1-\frac{\sigma_{lm}}{\tau_0}\right)+\frac{1}{2}\Delta(N_{l-1,m}+N_{l+1,m})=0\quad\text{для всех } l \text{ и } m, \qquad (18)$$

$$\left.\begin{aligned} B\delta_{l,0}-G_l&=0, \\ H_l&=0 \end{aligned}\right\}\quad\text{для всех } l. \qquad \begin{aligned}(19)\\(20)\end{aligned}$$

12.2.3. Выражения для амплитуд световых волн в дифрагировавших и отраженных спектрах. Перед тем как обсуждать решение уравнений (18)—(20), напишем выражения для полного светового возмущения в точке $(x,\ y,\ z)$, находящейся позади рассеивающей среды. Для этого подставим (5) в подынтег-

---

*) Символ Кронекера $\delta_{l,l'}$ нельзя спутать с параметром δ, введенным в (12.1.8), так как первый всегда снабжается индексами.

ральное выражение в правой части (3), проинтегрируем (3) по всему объему, занятому рассеивающей средой, и прибавим к полученному результату выражение для падающего поля. Мы вновь столкнемся с интегралом $\mathscr{Y}(\omega, p, q)$, рассматривавшимся в предыдущем разделе. Вспоминая, что в точке за рассеивающей средой $\mathscr{Y}(\omega, p, q)$ определяется (10б), и используя (18), получим для единственной не обращающейся в нуль компоненты полного прошедшего электрического поля следующее выражение:

$$E_z = \sum_l B_l \exp\left[(-i)\left\{\omega_l t - p_l x - [\omega_l^2 c^{-2} - p_l^2]^{1/2} y\right\}\right], \qquad (21)$$

где

$$B_l = \omega_l^2 \sum_m \sigma_{lm} N_{lm} \{\exp[ig_{lm} d]\} \{2c^2 g_{lm} [\omega_l^2 c^{-2} - p_l^2]^{1/2}\}^{-1}. \qquad (22)$$

Согласно (21) и (22) прошедшую волну можно считать состоящей из многих плоских волн с разными частотами и разными направлениями распространения. Подставляя (13) в экспоненту (21), легко получить выражения для частот ω_l и углов φ_l; они оказываются такими же, как и в п. 12.1.2.

Аналогично можно, воспользовавшись (10в), написать выражения для амплитуд $B_l^{(r)}$ в отраженном спектре, они имеют вид

$$B_l^{(r)} = -\omega_l^2 \sum_m \sigma_{lm} N_{lm} \{2c^2 h_{lm} [\omega_l^2 c^{-2} - p_l^2]^{1/2}\}^{-1}. \qquad (23)$$

Однако здесь мы рассматриваем только точки, находящиеся позади рассеивающей среды, так как в данной задаче интенсивности волн разных порядков в отраженном спектре, вообще говоря, очень малы.

12.2.4. Решение методом последовательных приближений. Предполагая, что амплитуда Δ ультразвуковой волны мала, мы решим в пп. 12.2.4 и 12.2.5 уравнения (18) — (20) и получим приближенные выражения для интенсивностей линий спектров первого и второго порядков в прошедшем свете. Случай, для которого это приближение не состоятельно, рассматривается качественно в п. 12.2.6; наконец, в п. 12.2.7 будут решены уравнения (18) — (20) в приближении, в основном эквивалентном приближению Рамана и Ната.

Сначала рассмотрим систему уравнений (18). Индекс m в этих уравнениях позволяет различать величины, относящиеся к различным допустимым значениям q. Следовательно, можно отбросить этот индекс и переписать (18) в виде

$$f_l(q^2) N_l(q^2) - \frac{1}{2}\Delta\left(N_{l-1}(q^2) + N_{l+1}(q^2)\right) = 0, \quad l = 0, \pm 1, \pm 2, \ldots, \qquad (24)$$

где

$$f_l(q^2) = \frac{\sigma_l(q)}{\tau_0} - 1 = \frac{3\,(p_l^2 + q^2 - n^2\omega_l^2 c^{-2})}{(n^2 - 1)(p_l^2 + q^2 + 2\omega_l^2 c^{-2})}. \qquad (25)$$

Уравнения (24) образуют бесконечную систему линейных однородных уравнений для амплитуд $N_l(q^2)$. Условием существования нетривиального решения, т. е. $N_l \neq 0$ для всех l, является равенство нулю детерминанта, образованного из коэффициентов при N_l. Корни этого уравнения, содержащего детерминант, определяют допустимые значения величин q^2. Обозначим их через q_m^2 $(m = 0, \pm 1, \pm 2, \ldots)$. Каждому такому q^2 соответствуют, конечно, два значения q, а именно $+|q|$ и $-|q|$, и два ряда амплитуд, т. е. $N_l(+|q|) \equiv N_l^+(q^2)$ и $N_l(-|q|) \equiv N_l^-(q^2)$ $(l = 0, \pm 1, \pm 2, \ldots)$. Теперь рекуррентное соотношение (24) определяет для данного допустимого значения q^2, например q_m^2, все значения $N_l^+(q_m^2)$ $(l = 0, \pm 1, \ldots)$ через одно из них, допустим $N_m^+(q_m^2)$. Так, все амплитуды $N_l^\pm(q^2)$ можно выразить через $N_m^\pm(q_m^2)$ $(m = 0, \pm 1, \pm 2, \ldots)$. Последние следует определить из уравнений (19) и (20), число которых как раз достаточно для этой цели (заметим, что в выражениях (5), (12), (15) и т. д. знак $\sum_m$ означает суммирование как по $N^+_{\ldots, m}$, так и по $N^-_{\ldots, m}$).

Получим теперь приближенное решение (24), используя метод возмущений. Следуя обычной процедуре метода возмущений и считая Δ малым параметром, разложим $\eta\,(=q^2)$ и N_l по степеням $^1/_2\,\Delta$; имеем

$$N_l(\eta) = N_l^{(0)} + \frac{1}{2}\Delta N_l^{(1)} + \left(\frac{1}{2}\Delta\right)^2 N_l^{(2)} + \ldots, \qquad (26а)$$

$$\eta = \eta^{(0)} + \frac{1}{2}\Delta\,\eta^{(1)} + \left(\frac{1}{2}\Delta\right)^2 \eta^{(2)} + \ldots \qquad (26б)$$

Используя (26б), можно записать $f_l(\eta)$ в виде

$$f_l(\eta) = f_l(\eta^{(0)}) + \frac{1}{2}\Delta\eta^{(1)} f_l'(\eta^{(0)}) + \left(\frac{1}{2}\Delta\right)^2 \left[\eta^{(2)} f_l'(\eta^{(0)}) + \frac{1}{2}(\eta^{(1)})^2 f_l''(\eta^{(0)})\right] + \ldots, \qquad (27)$$

где штрих у f означает дифференцирование по η. (Отметим, что $f_l'(\eta)$, $f_l''(\eta)$ ... не равны нулю для всех положительных вещественных значений η.) Подставляя (26) и (27) в соотношение (24), имеем ($l = 0, \pm 1, \pm 2, \ldots$)

$$\left\{f_l(\eta^{(0)}) + \frac{1}{2}\Delta\eta^{(1)} f_l'(\eta^{(0)}) + \left(\frac{1}{2}\Delta\right)^2 \left[\eta^{(2)} f_l'(\eta^{(0)}) + \frac{1}{2}(\eta^{(1)})^2 f_l''(\eta^{(0)})\right] + \ldots\right\} \times$$
$$\times \left\{N_l^{(0)} + \frac{1}{2}\Delta N_l^{(1)} + \left(\frac{1}{2}\Delta\right)^2 N_l^{(2)} + \ldots\right\} -$$
$$-\frac{1}{2}\Delta\left[N_{l-1}^{(0)} + \frac{1}{2}\Delta N_{l-1}^{(1)} + \ldots + N_{l+1}^{(0)} + \frac{1}{2}\Delta N_{l+1}^{(1)} + \ldots\right] = 0. \qquad (28)$$

Приравнивая сначала здесь нулю все члены, не зависящие от Δ, получим в нулевом приближении

$$f_l(\eta_l^{(0)})\, N_l^{(0)}(\eta) = 0 \qquad (l = 0, \pm 1, \pm 2, \ldots). \qquad (29)$$

Решениями уравнения (29) служат

$$\text{либо } f_l(\eta^{(0)}) = 0, \quad N_l^{(0)} \neq 0, \quad \text{либо } N_l^{(0)} = 0, \quad f_l(\eta^{(0)}) \neq 0. \qquad (30)$$

Обозначая через $\eta_l^{(0)}$ значение $\eta^{(0)}$, определяемое уравнением $f_l(\eta^{(0)}) = 0$, найдем из (30)

$$\eta_l^{(0)} = n^2\omega_l^2 c^{-2} - p_l^2, \quad N_l^{(0)}(\eta_l^{(0)}) \equiv N_{ll} \neq 0 \qquad (31)$$

и

$$N_{l'}^{(0)}(\eta_l^{(0)}) \equiv N_{l'l}^{(0)} = N_{ll}\delta_{l'l}. \qquad (32)$$

Далее, приравнивая нулю коэффициент при Δ в (28), получим

$$f_l(\eta_m^{(0)})\, N_l^{(1)}(\eta_m^{(0)}) + \eta_m^{(1)} f_l'(\eta_m^{(0)})\, N_l^{(0)}(\eta_m^{(0)}) = N_{l-1}^{(0)}(\eta_m^{(0)}) + N_{l+1}^{(0)}(\eta_m^{(0)}). \qquad (33)$$

Полагая здесь последовательно $l = m$, $m+1$, $m-1$, $m+2$, $m-2$, ... и используя (32), найдем

$$\eta_m^{(1)} = 0, \qquad (34а)$$

$$N_{m\pm 1,\, m}^{(1)} = \frac{N_{mm}}{f_{m\pm 1}(\eta_m^{(0)})} \qquad (34б)$$

и

$$N_{m\pm j,\, m}^{(1)} = 0 \quad \text{для} \quad j \geqslant 2. \qquad (34в)$$

Подобным же образом, приравнивая нулю коэффициент при Δ^2 в (28), получим следующие выражения для поправки к $\eta_m^{(0)}$ и для амплитуд N_{lm} до второго порядка *):

$$\eta_m^{(2)} = \frac{1}{f_m'(\eta_m^{(0)})}\left[\frac{1}{f_{m+1}(\eta_m^{(0)})} + \frac{1}{f_{m-1}(\eta_m^{(0)})}\right], \qquad (35а)$$

$$N_{m\pm 2,\, m}^{(2)} = \frac{N_{m,m}}{f_{m\pm 1}(\eta_m^{(0)})\, f_{m\pm 2}(\eta_m^{(0)})} \qquad (35б)$$

*) Приведенные здесь формулы справедливы, если все $f_{m\pm 1}(\eta_m^{(0)})$, $f_{m\pm 2}(\eta_m^{(0)})$,... отличны от нуля. Если это не так, то для учета вырождения указанный выше метод возмущений следует изменить. В расчетах возмущений до второго порядка вырождение сказывается при углах падения света $\theta = 0$ и $\theta = \arcsin(\lambda/2\Lambda)$. Однако здесь мы больше не будем заниматься этим и в п. 12.2.5 приведем окончательные результаты для данных двух случаев.

и

$$N^{(2)}_{m\pm 1,\, m}=0, \quad N^{(2)}_{m+l,\, m}=0 \quad \text{для} \quad l\geqslant 3. \tag{35в}$$

Как следует из предшествующих вычислений, в нулевом порядке теории возмущений отличны от нуля лишь величины N_{mm} ($m=0, \pm 1, \ldots$); в первом порядке отличны от нуля N_{mm} и $N_{m\pm 1\, m}$, а во втором порядке отличаются от нуля также и $N_{m\pm 2, m}$. Аналогично при вычислениях в более высоких порядках теории возмущений находим, что все большее число недиагональных амплитуд (т. е. амплитуд с разными индексами) отлично от нуля. Эти вычисления длинны, и мы не приводим их здесь. Однако можно предположить, что во всех случаях применимости теории возмущений можно пренебречь членами высших порядков.

Недиагональные амплитуды (34б) и (35б) полностью определены, если известны диагональные амплитуды, дающие решение (24) нулевого порядка. Мы находим последние из (19) и (20). Однако сначала поучительно исследовать решения этих уравнений в простом случае $\Delta=0$, для которого легко получить точное решение уравнений (18) — (20). Здесь возможными нулевыми амплитудами являются только диагональные амплитуды $N^{\pm}_{m,\, m}$ ($m=0, \pm 1, \pm 2, \ldots$). Если мы положим все недиагональные амплитуды в (19), (20) равными нулю, то найдем, что все амплитуды $N^{\pm}_{m,\, m}$ тождественно равны нулю, кроме $N^{\mp}_{0,\, 0}$ которые равны *)

$$N^{+}_{0,\,0}=\left(\frac{2B}{\sigma_{0,\,0}}\right)\cos\theta\,[(n^2-\sin^2\theta)^{1/2}-\cos\theta]\,(1+\rho^2-2\rho\cos 2q_0d)^{-1/2}\,\exp(i\psi),$$

$$N^{-}_{0,\,0}=N^{+}_{0,\,0}\left(\frac{g_{0,\,0}}{h_{0,\,0}}\right)\exp(2iq_0d),$$

где

$$\rho=\left|\frac{N^{-}_{0,\,0}}{N^{+}_{0,\,0}}\right|^2, \quad \psi=\operatorname{arctg}\left[\frac{\rho\sin 2q_0d}{1-\rho\cos 2q_0d}\right]. \tag{36}$$

Для нормального падения света ($\theta=0$) из соотношений (11), (31) и (36) находим

$$\rho_0=\left|\frac{N^{-}_{0,\,0}}{N^{+}_{0,\,0}}\right|^2_{\theta=0}=\frac{(n-1)^2}{(n+1)^2}. \tag{37}$$

С помощью (36), (27) и (23) легко получается следующее выражение для отражательной способности при нормальном падении на плоскопараллельную пластинку:

$$\left|\frac{B_0^{(r)}}{B}\right|^2=\frac{4\rho_0\sin^2 q_0d}{1+\rho_0^2-2\rho_0\cos 2q_0d},$$

что согласуется с (7.6.9), так как $q_0=nk$ при $\theta=0$.

Если Δ отличается от нуля, но остается еще достаточно малым, чтобы можно было применять метод возмущений, то из приведенного выше решения для $\Delta=0$ следует, что

$$N_{0,\,0}\gg N_{\pm 1,\,\pm 1}\gg N_{\pm 2,\,\pm 2}\ldots$$

Таким образом, для нахождения диагональных амплитуд нужно решить уравнения (19) и (20) методом последовательных приближений. Кроме того, поскольку для нормального или почти нормального падения света (в экспериментах по дифракции на ультразвуке θ равно самое большее 3°) отношение данного N^- к соответствующему N^+ мало (см (37)), можно вообще пренебречь

*) В выражениях для $N^{+}_{0,0}$ и $N^{-}_{0,0}$, приведенных в [15, 16], имеются опечатки. В нашем тексте они исправлены.

N^- и определять N^+ только из уравнений *) (19). Следует напомнить, что такая же аппроксимация подразумевается и в случае использования граничных условий (12.1.15).

Выражения для диагональных амплитуд $N^+_{0,0}$, $N^+_{\pm 1,\pm 1}$ и $N^+_{\pm 2,\pm 2}$ можно теперь получить из (19), используя (34б), (34в), (35б) и (35в). Далее выражения для интенсивностей линий в спектрах первого и второго порядков в прошедшем свете можно легко написать с помощью (22). Эти выражения будут приведены ниже.

12.2.5. Выражения для интенсивностей линий в спектрах первого и второго порядков в некоторых специальных случаях. *а) Случай* $\delta/\xi \ll 1$ *и* ξ *большого по сравнению с единицей.* Этот случай рассмотрен подробно выше. Интенсивности $I_{\pm 1}$ и $I_{\pm 2}$ линий в спектрах первого и второго порядков равны соответственно

$$I_{\pm 1} = |B_{\pm 1}|^2 = {}^1/_4 B^2\gamma^2\delta^2 \frac{\sin^2\left[\beta d\left(\xi \pm \frac{1}{2}\right)\right]}{\left(\xi \pm \frac{1}{2}\right)^2} \tag{38а}$$

и

$$I_{\pm 2} = |B_{\pm 2}|^2 = \frac{1}{64}B^2\gamma^4\delta^4 \left\{\frac{1}{2(\xi \pm 1)\left(\xi \pm \frac{1}{2}\right)\left(\xi \pm \frac{3}{2}\right)}\right\} \left[\frac{\sin^2\left[\beta d\left(\xi \pm \frac{1}{2}\right)\right]}{\left(\xi \pm \frac{1}{2}\right)} + \frac{\sin^2\left[\beta d\left(\xi \pm \frac{3}{2}\right)\right]}{\left(\xi \pm \frac{3}{2}\right)} - \frac{\sin^2[2\beta d(\xi \pm 1)]}{2(\xi \pm 1)}\right]. \tag{38б}$$

В этих уравнениях берутся либо все верхние, либо все нижние знаки.

Мы приведем без доказательств выражения для интенсивностей в двух других случаях, рассмотренных в работах [15, 16].

б) Случай $\xi \approx {}^1/_2$, $\delta \ll 1$

$$I_0 = \frac{1}{4}\frac{B^2\delta^2\gamma^2}{\left(\xi - \frac{1}{2}\right)^2 + \frac{1}{4}\delta^2\gamma^2}\left[\frac{\left(\xi - \frac{1}{2}\right)^2}{\frac{1}{4}\delta^2\gamma^2} + \cos^2\left[\beta d\left\{\left(\xi - \frac{1}{2}\right)^2 + \frac{1}{4}\delta^2\gamma^2\right\}^{1/2}\right]\right], \tag{39а}$$

$$I_{-1} = \frac{1}{4}\frac{B^2\delta^2\gamma^2}{\left(\xi - \frac{1}{2}\right)^2 + \frac{1}{4}\delta^2\gamma^2}\sin^2\left[\beta d\left\{\left(\xi - \frac{1}{2}\right)^2 + \frac{1}{4}\delta^2\gamma^2\right\}^{1/2}\right]. \tag{39б}$$

Выражения для I_1 и I_{-2} более сложны. Однако при $\xi = {}^1/_2$ они принимают простой вид

$$I_1 = \frac{1}{16}B^2\delta^2\gamma^2\left[-\sin^2\left(\frac{\pi d\Delta\gamma}{2\lambda n}\right) + 2\sum_{+,-}\sin^2\left\{\beta d\left(1 \pm \frac{1}{4}\delta\gamma\right)\right\}\right] \tag{39в}$$

и

$$I_{-2} = \frac{1}{16}B^2\delta^2\gamma^2\sin^2\left(\frac{\pi d\Delta\gamma}{2\lambda n}\right). \tag{39г}$$

в) Случай нормального падения света ($\xi = 0$), $\delta \ll 1$

$$I_1 = I_{-1} = B^2\delta^2\gamma^2\sin^2\left[\frac{1}{2}\beta d\left(1 + \frac{1}{8}\delta^2\gamma^2\right)\right], \tag{40а}$$

$$I_2 = I_{-2} = \frac{1}{48}B^2\delta^4\gamma^4\left[-\frac{1}{4}\sin^2 2\beta d + \sin^2\left\{\frac{1}{2}\beta d\left(1 + \frac{1}{8}\delta^2\gamma^2\right)\right\} + \frac{1}{3}\sin^2\left\{\frac{3}{2}\beta d\left(1 - \frac{1}{24}\delta^2\gamma^2\right)\right\}\right]. \tag{40б}$$

*) Это подробно разобрано в [15, 16]. В частности, там показано, что влияние N^- на амплитуды $B_l^{(r)}$ спектра в отраженном свете в общем случае нельзя считать пренебрежимо малым.

Пренебрегая величиной $\delta^2\gamma^2$, входящей в аргумент синуса в (40), можно также получить окончательные выражения из (38), полагая в них $\xi = 0$.

Как уже упоминалось, Бриллюэн [6] и Дэвид [7], а также Рытов [17] вывели выражения (38) для интенсивностей линий первого и второго порядков способом, описанным в § 12.1. Аггарвол [18] получил выражения (38) из дифференциальных уравнений (12.1.21) Рамана и Ната. Фаризо [19] показал, что выражения (39) для интенсивностей линий в первом и втором порядках при $\xi \approx \approx 1/2$ можно также получить из уравнений *) (21). В этих результатах, конечно, нет ничего удивительного, так как метод, основанный на дифференциальных уравнениях Максвелла, и использованный здесь метод интегральных уравнений эквивалентны.

12.2.6. Некоторые качественные результаты. Как ясно из приведенных выше выражений для интенсивности, при таких значениях δ и ξ, что либо $\delta \ll 1$, либо $\delta/\xi \ll 1$, с каждой стороны прошедшего пучка появляется лишь несколько спектров низших порядков и их интенсивности быстро убывают с увеличением порядка. Однако если ни одно из приведенных выше условий не выполняется, т. е. δ и δ/ξ велики по сравнению с единицей, то, вообще говоря, возникает значительно больше порядков. В этом случае решение уравнений (18) — (20), а значит, и расчет интенсивностей в разных порядках оказываются более трудными. Исследуя условия применимости метода возмущений, изложенного в п. 12.2.4, можно показать [15, 16], что при решении (18) — (20) надо считать амплитуды N_{lm} не равными нулю только тогда, когда значения обоих индексов l и m находятся между числами — M_1 и M_2, приближенно определяемыми соотношениями $(0 \leqslant \xi < \delta^{2/3})$

$$M_1 \approx \xi + \delta^{2/3} + 1, \qquad M_2 \approx -\xi + \delta^{2/3} + 1. \tag{41}$$

Так как Δ, вообще говоря, не может существенно превышать 10^{-4} и максимальное значение λ/Λ, при котором может наблюдаться явление дифракции, ограничивается практически достижимой разрешающей силой и т. д., то максимальное возможное значение δ $(= \Delta\Lambda^2/\lambda^2)$ примерно равно 100. Следовательно, даже в наиболее благоприятных экспериментальных условиях придется решать самое большее 20 совместных линейных уравнений из каждой бесконечной системы (18), (19), (20). Но даже при таком упрощении эти вычисления утомительны и их не производили.

Числа M_1 и M_2 указывают также число порядков, которые могут появиться с обеих сторон прямо прошедшего пучка света. Согласно (41) число порядков, возникающих по обе стороны прямо прошедшего пучка, должно становиться различным по мере того как $\xi = (\Lambda \sin \theta)/\lambda$ увеличивается от нуля. Причем большее число линий появляется с той стороны, куда отражается свет от волновых фронтов ультразвуковой волны. Партасарати [20] экспериментально

Таблица 12.1

Число наблюдающихся и предсказанных теорией порядков для различных углов падения θ при дифракции света на ультразвуковых волнах

θ	0	0°06′	0°22′	0°39′	1°01′	1°23′	1°45′	2°07′
ξ	0	0,6	2	4	6	8	10	13
M_1	5(5)	5(6)	6(7)	6(9)	3	2	1	1
M_2	5(5)	5(4)	3(3)	2(2)	2	1	1	1

*) Обсуждение экспериментальных результатов в связи с приведенными здесь различными выражениями для интенсивностей можно найти в статьях [15, 16].

изучал дифракционные спектры при разных углах падения θ; на рис. 12.3 показаны фотографии, взятые из его статьи. В этом эксперименте $(\lambda/\Lambda)=3\cdot10^{-3}$; полагая $\Delta\approx10^{-1}$, мы получим из (41) для нормального падения $M_1=M_2=5$. В табл. 12.1 указано число линий, фактически наблюдавшихся с каждой стороны прямо прошедшего пучка света при разных углах падения. В скобках приведены соответствующие теоретические значения, полученные из уравнения (41). (Конечно, если δ/ξ становится значительно меньше единицы, то с каждой стороны прямо прошедшего пучка появляются только один или два порядка.)

Рис. 12.3. Дифракция света на ультразвуковых волнах [20]

Спектры наблюдались при различных углах падения θ ($\lambda/\Lambda=3\cdot10^{-3}$, $\Delta\approx10^{-4}$).

Аналогичные экспериментальные результаты были получены также Номото [21]. Зависимости M_1 и M_2 от θ, полученные в его работе, хотя и находятся в качественном согласии с (41), показывают также небольшое периодическое изменение с ξ в области $0\leqslant\xi<\delta^{3/2}$. Уравнения (41) основываются на слишком грубых соображениях, чтобы объяснить эту особенность.

12.2.7. Приближение Рамана — Ната. Покажем теперь, что полученные Раманом и Натом выражения для интенсивностей, содержащие функции Бесселя, можно получить и из решения уравнений (18)—(20). Пренебрегая небольшим изменением частот ω_l в зависимости от l и вспоминая, что $q_m^2\approx(n^2\omega_m^2c^{-2}-p_m^2)$, можно с хорошим приближением принять в знаменателе (25) $p_l^2+q^2+2\omega_l^2c^{-2}\approx k^2(n^2+2)$. Поэтому (24) можно представить в виде

$$\frac{q^2-n^2k^2+k^2\sin^2\theta}{\frac{1}{2}\Delta\gamma k^2}N_l(q)-N_{l+1}(q)-N_{l-1}(q)=$$
$$=-\frac{l^2+2l\xi}{2\gamma\delta}N_l(q)\quad(l=0,\ \pm1,\ \pm2,\ \ldots).\tag{42}$$

Выражения для интенсивностей, содержащие функции Бесселя, получены Раманом и Натом при пренебрежении членами $[(l^2+2l\xi)/\delta]U_l$ в уравнениях (12.1.21). В том же приближении мы можем пренебречь правой частью у (42). Тогда (42) можно переписать следующим образом:

$$\frac{q^2-b^2}{\mu^2}N_l(q)=N_{l+1}(q)+N_{l-1}(q)\qquad(l=0,\ \pm1,\ \ldots),\tag{43}$$

где

$$b^2=k^2(n^2-\sin^2\theta)\quad\text{и}\quad\mu^2=\frac{1}{2}\Delta\gamma k^2.\tag{44}$$

Для решения *) (43) допустим, что $N_l(q)$ имеет вид

$$N_l(q)=N\exp\left(\frac{2\pi ilm}{M}\right),\tag{45}$$

где M — некоторое очень большое целое число, а m — такое целое число, что $0\leqslant m<M$. (Мы сейчас покажем, что окончательный результат не зависит от

*) Уравнения (43) аналогичны соотношениям, определяющим нормальные типы колебаний линейной цепочки атомов (см. [22]).

M.) Подставляя (45) в (43), получим для допустимых значений q^2 соотношение

$$q_m^2 = b^2 + 2\mu^2 \cos\left(\frac{2\pi m}{M}\right), \quad 0 \leqslant m < M, \tag{46}$$

или, вспоминая, что $\mu^2 \ll b^2$,

$$\pm|q_m| = \pm\left[b + \left(\frac{\mu^2}{b}\right)\cos\left(\frac{2\pi m}{M}\right)\right]. \tag{47}$$

Следовательно, для каждого m существуют два значения q, а именно $\pm|q_m|$; соответствующие амплитуды получаются из выражения

$$N_l(\pm q_m) \equiv N_{lm}^{\pm} = N^{\pm} \exp\left(\frac{2\pi i l m}{M}\right). \tag{48}$$

Обе постоянные N^+ и N^- в (48) теперь можно найти из (19) и (20). Как и в п. 12.2.4, мы пренебрегаем амплитудами N_{lm}^- и определяем постоянную N^+ только из (19). Заметим, что при не слишком больших значениях l величина $[\sigma_l(q)/g_l(q)] \approx [\sigma_0(q_0^{(0)})/g_0(q_0^{(0)})]$; поэтому можно показать, что уравнения (19) тождественно удовлетворяются, если принять в (48)

$$N^+ = \frac{1}{M} 2Bc^2\omega^{-2}[\omega^2 c^{-2} - p_0^2]^{1/2}\left(\frac{g_{0,0}}{\sigma_{0,0}}\right). \tag{49}$$

Подставляя (48) и (49) в (22) и полагая $(\omega_l^2 c^{-2} - p_l^2) = (\omega^2 c^{-2} - p_0^2)$, получим выражение для амплитуды B_l дифрагировавшей волны l-го порядка

$$B_l \approx \frac{B}{M}\sum_{m=0}^{M-1} \exp\left\{i\left[\frac{2\pi l m}{M} + \left(\frac{\mu^2 d}{b}\right)\cos\left(\frac{2\pi m}{M}\right)\right]\right\}. \tag{50}$$

Здесь фазовый множитель, не зависимый от m, опущен. Теперь, поскольку мы предположили, что M — очень большое целое число, можно заменить ряд в (50) интегралом. Полагая $2\pi m/M = \psi'$ и $d\psi' = 2\pi/M$, перепишем (50) в виде

$$B_l \approx \frac{B}{2\pi}\int_0^{2\pi} \exp\left\{i\left[l\psi' + \left(\frac{\mu^2 d}{b}\right)\cos\psi'\right]\right\} d\psi', \tag{51}$$

который не зависит от M. Разделяя интеграл в (51) на две части: одну от 0 до $3\pi/2$ и другую от $3\pi/2$ до 2π — и полагая в них $\psi' = \pi/2 - \psi$ и $\psi' = 5\pi/2 - \psi$ соответственно, получим [23]

$$B_l \approx \frac{B}{2\pi}\int_{-\pi}^{\pi} \exp\left\{i\left[\frac{1}{2}l\pi - l\psi + \left(\frac{\mu^2 d}{b}\right)\sin\psi\right]\right\} d\psi = B\exp\left(\frac{1}{2}il\pi\right) J_l\left(\frac{\mu^2 d}{b}\right). \tag{52}$$

Следовательно, интенсивности $I_l = |B_l|^2$ точно равны $B^2 J_l^2(\mu^2 d/b)$. Кроме того, с помощью (44), (12.1.5) и (12.1.8) можно показать, что аргумент $(\mu^2 d/b)$ функций Бесселя J_l такой же, как и в выражениях, полученных Раманом и Натом и приведенных на стр. 553.

ЛИТЕРАТУРА

1. Bergmann, Der Ultraschall, Hirzel, Zürich, 1954. (Л. Бергман, Ультразвук и его применение в науке и технике, ИЛ, 1957.)
2. L. Brillouin, Ann. de Physique 17, 103 (1921).
3. P. Debye, F. W. Sears, Proc. Nat. Acad. Sci. USA 18, 409 (1932).
4. R. Lucas, P. Biquard, J. Phys. Radium 3, 464 (1932).
5. G. W. Willard, J. Acoust. Soc. Amer. 21, 101 (1949).
6. L. Brillouin, La Diffraction de la Lumière par des Ultrasons Hermann, Paris, 1933.
7. E. David, Phys. Z. 38, 587 (1937).
8. C. V. Raman, N. S. N. Nath, Proc. Indian Acad. Sci. A2, 406, 413 (1935); 3, 75, 119 (1936).
9. Rayleigh, Proc. Roy. Soc. A79, 399 (1907).

10. E. T. Whittaker, G. N. Watson, A Course of Modern Analysis, 4th ed. Cambr. Univ. Press. 1946. (Э. Уиттекер, Г. Ватсон, Курс современного анализа, Гостехиздат, 1933.)
11. R. Extermann, G. Wannier, Helv Phys. Acta 9, 520 (1936).
12. N. S. N. Nath, Proc. Indian Acad Sci A4, 222 (1936), A8, 499 (1938).
13. R. R. Aggarwal, Ph. D. Thesis, Delhi University, India, 1954
14. E. H. Wagner, Z f. Phys. 141, 604, 622 (1955)
15. W. J. Noble, Ph D. Thesis, University of Edinburgh, 1952.
16. A. B. Bhatia, W. J. Noble, Proc Roy. Soc. A220, 356, 369 (1953).
17. С. М. Рытов, Известия АН СССР, сер физ, № 2, 223 (1937).
18. R. R Aggarwal, Proc. Indian Acad Sci A31, 417 (1950).
19. P. Phariseau, Proc. Indian Acad Sci. A44, 165 (1956).
20. S. Parthasarathy, Proc. Indian Acad. Sci. A3, 442 (1936).
21. O. Nomoto, Proc. Phys-Math. Soc. Japan 24, 380, 613 (1942).
22. M. Born, Th. v. Kármán, Phys. Z. 13, 297 (1912).
23. H. Jeffreys, B. S. Jeffreys, Methods of Mathematical Physics, Cambr. Univ. Press, 1946, p. 547.

ГЛАВА 13

МЕТАЛЛООПТИКА

До сих пор мы рассматривали распространение света в непроводящих изотропных средах. Теперь обратимся к оптике проводящих сред, главным образом металлов. Обычный кусок металла состоит из небольших кристаллов, ориентированных случайным образом. Монокристаллы заметных размеров встречаются редко, но их можно приготовить в лаборатории. Оптические свойства кристаллов рассматриваются в гл. 14. Очевидно, что совокупность случайным образом ориентированных кристаллов ведет себя как изотропное тело, а поскольку в проводящей изотропной среде теория распространения света значительно проще, чем в кристалле, мы довольно подробно рассмотрим ее здесь.

Согласно § 1.1 проводимость связана с выделением джоулева тепла. Это — необратимое явление, при котором электромагнитная энергия исчезает или, точнее, превращается в тепло, в результате чего электромагнитная волна в проводнике затухает. Вследствие чрезвычайно высокой проводимости металлов этот эффект в них столь велик, что они практически непрозрачны. Указанное свойство позволяет металлам играть важную роль в оптике. Сильное поглощение сопровождается высокой отражательной способностью, так что металлические поверхности служат прекрасными зеркалами. Частичное проникновение света в металл (хотя глубина проникновения и мала) дает возможность получать информацию о константах металлов и механизме поглощения из наблюдений отраженного света.

Вначале мы чисто формально рассмотрим результаты, вытекающие из наличия проводимости, а затем кратко обсудим простую, до некоторой степени идеализированную физическую модель этого явления, основанную на классической электронной теории. Такая модель дает лишь грубое объяснение некоторым из наблюдающихся эффектов; более точную модель можно создать лишь с помощью квантовой механики, однако это выходит за рамки настоящей книги. Формальную теорию мы применим к двум проблемам, представляющим практический интерес: к оптике слоистых сред, содержащих поглощающий элемент, и к дифракции света на металлической сфере *).

Чрезвычайно привлекательной математической особенностью теории является то, что наличие проводимости можно учесть, просто вводя вместо вещественной диэлектрической проницаемости комплексную (или комплексный показатель преломления). В металлах преобладает мнимая ее часть.

§ 13.1. Распространение волн в проводнике

Рассмотрим однородную изотропную среду с диэлектрической проницаемостью ε, магнитной проницаемостью μ и проводимостью σ. Используя материальные уравнения (1.1.9) — (1.1.11), а именно $\mathbf{j}=\sigma\mathbf{E}$, $\mathbf{D}=\varepsilon\mathbf{E}$, $\mathbf{B}=\mu\mathbf{H}$,

) На русском языке вопросы металлооптики несколько в другом аспекте изложены в обзорных статьях Гинзбурга и Мотулевич [81] и Мотулевич [81а*] и в монографии Соколова [82*]. (*Прим. перев.*)

запишем уравнения Максвелла в виде

$$\operatorname{rot} \mathbf{H} - \frac{\varepsilon}{c}\dot{\mathbf{E}} = \frac{4\pi}{c}\sigma\mathbf{E}, \tag{1}$$

$$\operatorname{rot} \mathbf{E} + \frac{\mu}{c}\dot{\mathbf{H}} = 0, \tag{2}$$

$$\operatorname{div} \mathbf{E} = \frac{4\pi}{\varepsilon}\rho, \tag{3}$$

$$\operatorname{div} \mathbf{H} = 0. \tag{4}$$

Легко видеть, что для электромагнитного возмущения, падающего извне на проводник, мы можем заменить (3) уравнением $\operatorname{div} \mathbf{E} = 0$. Действительно, если мы применим операцию дивергенции к уравнению (1) и используем (3), то получим

$$-\frac{\varepsilon}{c}\operatorname{div}\dot{\mathbf{E}} = \frac{4\pi}{c}\sigma\frac{4\pi}{\varepsilon}\rho.$$

Дифференцируя уравнение (3) по времени, найдем

$$\operatorname{div}\dot{\mathbf{E}} = \frac{4\pi}{\varepsilon}\dot{\rho}.$$

Исключая $\operatorname{div}\dot{\mathbf{E}}$ из двух последних уравнений, получим

$$\dot{\rho} + \frac{4\pi\sigma}{\varepsilon}\rho = 0, \tag{5}$$

или после интегрирования

$$\rho = \rho_0 \exp\left(-\frac{t}{\tau}\right), \quad \text{где} \quad \tau = \frac{\varepsilon}{4\pi\sigma}. \tag{6}$$

Таким образом, видно, что любая плотность электрического заряда ρ экспоненциально уменьшается со временем. *Время релаксации* τ чрезвычайно мало для любой среды, обладающей заметной проводимостью. Для металлов это время значительно меньше периода колебаний волны; например, для света в оранжевой области видимого спектра период колебаний равен $2\cdot 10^{-15}$ *сек*, тогда как для меди τ/ε порядка $2\cdot 10^{-19}$ *сек*. Для любого разумного значения ε, которого можно ожидать, τ так мало по сравнению с периодом световой волны, что в металле ρ всегда практически равно нулю. Тогда уравнение (3) можно переписать в виде

$$\operatorname{div} \mathbf{E} = 0. \tag{7}$$

Из (1) и (2) после исключения $\mathbf{H}$ и использования (7) следует, что $\mathbf{E}$ удовлетворяет волновому уравнению

$$\nabla^2\mathbf{E} = \frac{\mu\varepsilon}{c^2}\ddot{\mathbf{E}} + \frac{4\pi\mu\sigma}{c^2}\dot{\mathbf{E}}. \tag{8}$$

Наличие члена с $\dot{\mathbf{E}}$ означает затухание волны, т. е. при распространении через среду волна постепенно ослабевает.

Если поле строго монохроматично и обладает циклической частотой ω, т. е. если $\mathbf{E}$ и $\mathbf{H}$ имеют вид $\mathbf{E} = \mathbf{E}_0\exp(-i\omega t)$, то производная $\partial/\partial t = -i\omega$, и уравнения (1) и (2) можно переписать следующим образом:

$$\operatorname{rot} \mathbf{H} + \frac{i\omega}{c}\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right)\mathbf{E} = 0, \tag{9}$$

$$\operatorname{rot} \mathbf{E} - \frac{i\omega\mu}{c}\mathbf{H} = 0. \tag{10}$$

Тогда уравнение (8) примет вид

$$\nabla^2\mathbf{E} + \hat{k}^2\mathbf{E} = 0, \tag{11}$$

где

$$\hat{k}^2 = \frac{\omega^2\mu}{c^2}\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right). \tag{12}$$

Если в эти уравнения ввести величину

$$\hat{\varepsilon} = \varepsilon + i\frac{4\pi\sigma}{\omega}, \tag{13}$$

то они формально станут идентичными с соответствующими уравнениями для непроводящих сред, где фигурирует вещественная диэлектрическая проницаемость ε.

Аналогия с непроводящими средами станет еще ближе, если, кроме комплексного волнового числа $\hat{k}$ и комплексной диэлектрической проницаемости $\hat{\varepsilon}$, ввести также комплексную фазовую скорость $\hat{v}$ и комплексный показатель преломления $\hat{n}$, которые по аналогии с (1.2.8), (1.2.12) и (1.3.21) определяются как

$$\hat{v} = \frac{c}{\sqrt{\mu\hat{\varepsilon}}}, \qquad \hat{n} = \frac{c}{\hat{v}} = \sqrt{\mu\hat{\varepsilon}} = \frac{c}{\omega}\hat{k}. \tag{14}$$

Положим

$$\hat{n} = n(1 + i\kappa), \tag{15}$$

где n и κ вещественны, и назовем κ *показателем затухания* *). Величины n и κ легко выразить через материальные постоянные ε, μ и σ. Возводя в квадрат обе части (15), получим

$$\hat{n}^2 = n^2(1 + 2i\kappa - \kappa^2).$$

Кроме того, из (14) и (13) имеем

$$\hat{n}^2 = \mu\hat{\varepsilon} = \mu\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right).$$

Приравнивая вещественные и мнимые части в этих двух выражениях для $\hat{n}^2$, получим

$$n^2(1 - \kappa^2) = \mu\varepsilon, \tag{16а}$$

$$n^2\kappa = \frac{2\pi\mu\sigma}{\omega} = \frac{\mu\sigma}{\nu}. \tag{16б}$$

Отсюда следует, что

$$n^2 = \frac{1}{2}\left\{\sqrt{\mu^2\varepsilon^2 + \frac{4\mu^2\sigma^2}{\nu^2}} + \mu\varepsilon\right\}, \tag{17а}$$

$$n^2\kappa^2 = \frac{1}{2}\left\{\sqrt{\mu^2\varepsilon^2 + \frac{4\mu^2\sigma^2}{\nu^2}} - \mu\varepsilon\right\}. \tag{17б}$$

Здесь выбран положительный квадратный корень, так как n и $n\kappa$ вещественны, а следовательно, n^2 и $n^2\kappa^2$ должны быть положительными.

Уравнение (11) формально идентично волновому уравнению для непроводящей среды, но теперь волновое число комплексно. Простейшим решением (11) служит плоская гармоническая во времени волна

$$\mathbf{E} = \mathbf{E}_0 \exp\{i[\hat{k}(\mathbf{r}\cdot\mathbf{s}) - \omega t]\}. \tag{18}$$

Если в соответствии с (14) и (15) подставить значение $\hat{k}$ из соотношения $\hat{k} = \omega\hat{n}/c = \omega n(1 + i\kappa)/c$, то (18) примет вид

$$\mathbf{E} = \mathbf{E}_0 \exp\left[-\frac{\omega}{c}n\kappa(\mathbf{r}\cdot\mathbf{s})\right]\exp\left\{i\omega\left[\frac{n}{c}(\mathbf{r}\cdot\mathbf{s}) - t\right]\right\}.$$

Вещественная часть этого выражения, а именно

$$\mathbf{E} = \mathbf{E}_0 \exp\left[-\frac{\omega}{c}n\kappa(\mathbf{r}\cdot\mathbf{s})\right]\cos\left\{\omega\left[\frac{n}{c}(\mathbf{r}\cdot\mathbf{s}) - t\right]\right\}, \tag{19}$$

*) Употребляется также термин «коэффициент экстинкции».

представляющая собой электрический вектор, является плоской волной длиной $\lambda = 2\pi c/\omega n$, затухание которой определяется экспоненциальным членом. Так как плотность энергии w волны пропорциональна среднему по времени от $\mathbf{E}^2$, то ясно, что w будет уменьшаться в соответствии с законом

$$w = w_0 \exp[-\chi(\mathbf{r}\cdot\mathbf{s})], \tag{20}$$

где

$$\chi = \frac{2\omega}{c} n\kappa = \frac{4\pi\nu}{c} n\kappa = \frac{4\pi}{\lambda_0} n\kappa = \frac{4\pi}{\lambda}\kappa. \tag{21}$$

Здесь λ_0 — длина волны в вакууме, λ — длина волны в среде. Постоянная χ называется *коэффициентом поглощения*.

Плотность энергии падает в e раз на расстоянии d, где

$$d = \frac{1}{\chi} = \frac{\lambda_0}{4\pi n\kappa} = \frac{\lambda}{4\pi\kappa}. \tag{22}$$

Обычно эта величина составляет очень малую долю длины волны (см. табл. 13.1) *).

Таблица 13.1

«Глубина проникновения» d излучения с различными длинами волн λ_0 для меди ($\sigma \approx 5{,}14\cdot 10^{17}$ *сек*$^{-1}$, $\mu = 1$)

	Ультрафиолетовая область	Инфракрасная область	СВЧ	Длинные радиоволны
λ_0	$1000 Å = 10^{-5}$ *см*	10 *мк* $= 10^{-3}$ *см*	10 *см*	1000 *м* $= 10^5$ *см*
d	$6{,}2\cdot 10^{-8}$ *см*	$6{,}2\cdot 10^{-7}$ *см*	$6{,}2\cdot 10^{-5}$ *см*	$6{,}2\cdot 10^{-3}$ *см*

Возвращаясь к уравнениям (17), мы видим, что когда $\sigma = 0$, первое уравнение точно переходит в соотношение Максвелла (1.2.14) $n^2 = \mu\varepsilon$, а второе дает $\kappa = 0$. Для металлов $\sigma \neq 0$ и фактически так велико, что в (17) можно пренебречь величиной ε по сравнению с $2\sigma/\nu$. Чтобы оценить порядок рассматриваемых величин, заметим, что для большинства металлов **) $\sigma \sim 10^{17}$ *сек*$^{-1}$, и поэтому для $\lambda = 5500$ А (что соответствует частоте $\nu = 5{,}5\cdot 10^{14}$ *сек*$^{-1}$) $\sigma/\nu \sim 200$. Диэлектрическую проницаемость ε металла невозможно измерить прямо, но как мы покажем, ее легко получить из оптических экспериментов. Однако механизмы электрической поляризации в металлах и диэлектриках не могут фундаментально отличаться друг от друга, и поэтому мы вправе предположить, что в обоих типах сред ε одинаково по порядку величины. Следовательно, для не слишком коротких длин волн можно считать, что

$$\frac{\mu\sigma}{\nu} = n^2\kappa \gg \mu\varepsilon. \tag{23}$$

Уравнения (17) и (22) переходят при этом в

$$n \approx n\kappa = \sqrt{\frac{\mu\sigma}{\nu}}, \tag{24}$$

$$d \approx \frac{\lambda_0}{4\pi}\sqrt{\frac{\nu}{\mu\sigma}} = \frac{1}{4\pi}\sqrt{\frac{c\lambda_0}{\mu\sigma}}. \tag{25}$$

Идеальный проводник характеризуется бесконечной проводимостью ($\sigma \to \infty$). Так как, согласно (16), $\varepsilon/\sigma = (1-\kappa^2)/\nu\kappa$, то в этом предельном случае

*) Явление проникновения на глубину, составляющую малую часть длины волны, хорошо известно для переменных токов. В технике оно называется «скин-эффектом».

**) Значение, используемое авторами для величин σ, относится к *статической* проводимости. Однако в формуле (17) должна фигурировать проводимость *на частоте* ν. Эта величина в оптическом диапазоне для типичных металлов порядка 10^{15} *сек*$^{-1}$. (*Прим. перев.*)

$\kappa^2 \to 1$ или на основании (16а) $n \to \infty$. Такой проводник вообще не позволял бы электромагнитной волне проникать на какую-либо глубину и отражал бы весь падающий свет (см. ниже, § 13.2).

Если показатель преломления прозрачных веществ можно легко измерить по углу преломления, то такие измерения для металлов исключительно трудны, так как металлический образец, пропускающий заметную долю падающего света, должен быть чрезвычайно тонким. Тем не менее Кундту [1] удалось изготовить металлические призмы и провести прямые измерения вещественной и мнимой частей комплексного показателя преломления. Однако обычно оптические постоянные металлов определяются посредством катоптрических, а не диоптрических экспериментов, т. е. путем изучения тех изменений, которые возникают при отражении света от металла, а не при прохождении через него.

§ 13.2. Преломление и отражение на поверхности металла

Мы видели, что основные уравнения, описывающие распространение плоской гармонической волны в проводящей среде, отличаются от соответствующих уравнений для прозрачного диэлектрика лишь тем, что вещественные постоянные ε и k заменяются на комплексные $\hat{\varepsilon}$ и $\hat{k}$. Отсюда следует, что формулы, полученные в гл. 1, применимы и в данном случае, поскольку они содержат лишь линейные соотношения между компонентами векторов поля плоских монохроматических волн. В частности, остаются справедливыми граничные условия для распространения волны через поверхность раздела, а, следовательно, также формулы, приведенные в § 1.5 и относящиеся к преломлению и отражению.

Рассмотрим вначале распространение плоской волны из диэлектрика в проводник, причем обе среды будем считать бесконечными, а за поверхность раздела между ними выберем плоскость $z=0$. По аналогии с (1.5.8) закон преломления можно записать в виде

$$\sin\theta_t = \frac{1}{\hat{n}}\sin\theta_i. \tag{1}$$

Так как $\hat{n}$ — комплексно, то комплексным окажется и θ_t. Таким образом, эта величина уже не имеет простого смысла угла преломления.

Выберем в качестве плоскости падения плоскость xz. Тогда зависящая от координат часть фазы волны в проводнике определится выражением $\hat{k}(\mathbf{r}\cdot\mathbf{s}^{(t)})$, где (см. (1.5.4))

$$s_x^{(t)} = \sin\theta_t, \qquad s_y^{(t)} = 0, \qquad s_z^{(t)} = \cos\theta_t. \tag{2}$$

Из (1) и (2) и (13.1.15) имеем

$$s_x^{(t)} = \sin\theta_t = \frac{\sin\theta_i}{n(1+i\kappa)} = \frac{1-i\kappa}{n(1+\kappa^2)}\sin\theta_i, \tag{3а}$$

$$s_z^{(t)} = \cos\theta_t = \sqrt{1-\sin^2\theta_t} = \sqrt{1-\frac{(1-\kappa^2)}{n^2(1+\kappa^2)^2}\sin^2\theta_i + i\,\frac{2\kappa}{n^2(1+\kappa^2)^2}\sin^2\theta_i}. \tag{3б}$$

Удобно выразить $s_z^{(t)}$ в форме

$$s_z^{(t)} = \cos\theta_t = qe^{i\gamma} \tag{4}$$

(q, γ вещественны). Можно сразу же выразить q и γ через n, κ и $\sin\theta_i$, если возвести в квадрат соотношения (3б) и (4) и приравнять отдельно вещественные и мнимые части.

Это дает

$$q^2\cos 2\gamma = 1 - \frac{1-\kappa^2}{n^2(1+\kappa^2)^2}\sin^2\theta_i, \qquad q^2\sin 2\gamma = \frac{2\kappa}{n^2(1+\kappa^2)^2}\sin^2\theta_i. \tag{5}$$

Отсюда следует, что

$$\hat{k}(\mathbf{r}\cdot\mathbf{s}^{(t)})=\frac{\omega}{c}n(1+i\kappa)(xs_x^{(t)}+zs_z^{(t)})=$$
$$=\frac{\omega}{c}n(1+i\kappa)\left[\frac{x(1-i\kappa)}{n(1+\kappa^2)}\sin\theta_i+z(q\cos\gamma+iq\sin\gamma)\right]=$$
$$=\frac{\omega}{c}[x\sin\theta_i+znq(\cos\gamma-\kappa\sin\gamma)+inzq(\kappa\cos\gamma+\sin\gamma)]. \quad (6)$$

Мы видим, что поверхности постоянной амплитуды определяются уравнением

$$z=\text{const} \quad (7)$$

и поэтому являются плоскостями, параллельными поверхности раздела. Поверхности постоянной вещественной фазы определяются уравнением

$$x\sin\theta_i+znq(\cos\gamma-\kappa\sin\gamma)=\text{const} \quad (8)$$

и являются плоскостями, нормали которых образуют угол θ_t' с нормалью к границе, причем

$$\left.\begin{aligned}\cos\theta_t'&=\frac{nq(\cos\gamma-\kappa\sin\gamma)}{\sqrt{\sin^2\theta_i+n^2q^2(\cos\gamma-\kappa\sin\gamma)^2}},\\ \sin\theta_t'&=\frac{\sin\theta_i}{\sqrt{\sin^2\theta_i+n^2q^2(\cos\gamma-\kappa\sin\gamma)^2}}.\end{aligned}\right\} \quad (9)$$

В общем случае поверхности постоянной амплитуды и поверхности постоянной фазы не совпадают друг с другом, и поэтому волна в металле оказывается *неоднородной*.

Если квадратный корень в (9) обозначить через n', то уравнение для $\sin\theta_t'$ можно переписать в виде $\sin\theta_t'=\sin\theta_i/n'$, т. е. в форме закона Снеллиуса. Однако здесь n' зависит не только от величин, характеризующих среду, но и от угла падения θ_i.

Подставляя комплексное значение для θ_t из (1) в формулы Френеля (см. п. 1.5 2), легко также получить выражения для амплитуд и фаз преломленной и отраженной волн В явном виде они будут приведены в п 13.4.1 при изложении теории слоистых проводящих сред. Здесь мы покажем, как можно получить оптические постоянные металла из наблюдений отраженной волны.

Так как, по предположению, первой средой служит диэлектрик, то отраженная волна будет обычной (однородной) волной с вещественной фазой. Как и в (1.5.21а), компоненты амплитуды падающей волны $A_\parallel$ и $A_\perp$ и соответствующие компоненты отраженной волны $R_\parallel$ и $R_\perp$ связаны соотношениями

$$R_\parallel=\frac{\operatorname{tg}(\theta_i-\theta_t)}{\operatorname{tg}(\theta_i+\theta_t)}A_\parallel,\quad R_\perp=-\frac{\sin(\theta_i-\theta_t)}{\sin(\theta_i+\theta_t)}A_\perp. \quad (10)$$

Поскольку угол θ_t комплексный, то комплексны и отношения $R_\parallel/A_\parallel$ и $R_\perp/A_\perp$, т е при отражении происходят характерные изменения фазы Таким образом, падающий линейно поляризованный свет при отражении от поверхности металла в общем случае становится эллиптически поляризованным

Пусть $\varphi_\parallel$ и $\varphi_\perp$ — фазы, а $\rho_\parallel$ и $\rho_\perp$ — абсолютные значения коэффициентов отражения, т. е.

$$r_\parallel=\frac{R_\parallel}{A_\parallel}=\rho_\parallel\exp(i\varphi_\parallel),\quad r_\perp=\frac{R_\perp}{A_\perp}=\rho_\perp\exp(i\varphi_\perp). \quad (11)$$

Предположим, что падающий свет *линейно поляризован* и азимутальный угол равен α_i, т. е.

$$\operatorname{tg}\alpha_i=\frac{A_\perp}{A_\parallel}, \quad (12)$$

и пусть α_r — азимутальный угол (обычно комплексный) для отраженного

света. Тогда *)

$$\operatorname{tg}\alpha_r = \frac{R_\perp}{R_\parallel} = -\frac{\cos(\theta_i - \theta_t)}{\cos(\theta_i + \theta_t)} \operatorname{tg}\alpha_i = Pe^{-i\Delta} \operatorname{tg}\alpha_i, \tag{13}$$

где

$$P = \frac{\rho_\perp}{\rho_\parallel}, \qquad \Delta = \varphi_\parallel - \varphi_\perp. \tag{14}$$

Заметим, что величина α_r вещественна в следующих двух случаях:

1. При нормальном падении ($\theta_i = 0$); тогда $P = 1$ и $\Delta = -\pi$, так что $\operatorname{tg}\alpha_r = -\operatorname{tg}\alpha_i$.

2. При скользящем падении ($\theta_i = \pi/2$); тогда $P = 1$ и $\Delta = 0$, так что $\operatorname{tg}\alpha_r = \operatorname{tg}\alpha_i$.

Необходимо напомнить, что в случае нормального падения направления падающего и отраженного лучей противоположны, таким образом, отрицательная величина $\operatorname{tg}\alpha_r$ означает, что направление поляризации линейно поляризованного света не изменяется в пространстве. Оно не меняется и при скользящем угле падения.

Между двумя только что рассмотренными экстремальными углами существует так называемый *главный угол падения* $\bar{\theta}_i$, для которого $\Delta = -\pi/2$. При этом угле линейно поляризованный свет в общем случае превращается после отражения в эллиптически поляризованный свет, но, как легко видеть из (1.4 31б) (для $\delta = \pi/2$), одна из осей эллипса поляризации параллельна, а другая перпендикулярна к плоскости падения Если, кроме того, $P \operatorname{tg}\alpha_i = 1$, то, согласно (13), $\operatorname{tg}\alpha_r = -i$ и отраженный свет поляризован *по кругу*.

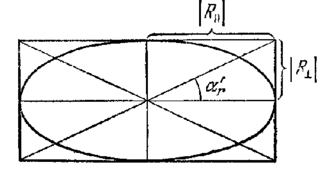


Рис 13 1 Эллипс поляризации для света, отраженного от металла при главном угле падения.

Предположим, что на металл падает линейно поляризованный свет, и между $R_\parallel$ и $R_\perp$ вводится с помощью подходящего компенсатора (см п 14.4.2) дополнительная разность фаз Δ. Если полная разность фаз равна нулю, то, согласно (13), отраженный свет линейно поляризован с азимутом α_r', определяемым из соотношения

$$\operatorname{tg}\alpha_r' = P \operatorname{tg}\alpha_i. \tag{15}$$

По очевидным причинам угол α_r' называется *углом восстановленной поляризации*, хотя его обычно определяют только для падающего света, линейно поляризованного с азимутом **) $\alpha_i = 45°$. Значения α_r' и P, относящиеся к главному углу падения $\theta_i = \bar{\theta}_i$, обозначим соответственно через $\bar{\alpha}_r'$ и $\bar{P}$. Если мы представим, что вокруг эллипса поляризации отраженного света, падающего под главным углом (дополнительная компенсация отсутствует), описан прямоугольник, стороны которого параллельны и перпендикулярны плоскости падения, то отношение его сторон составляет $\bar{P} \operatorname{tg}\alpha_i$, а угол между диагональю и плоскостью падения равен $\bar{\alpha}_r'$ (рис. 13.1).

Для дальнейшего полезно ввести такой угол ψ, что

$$\operatorname{tg}\psi = P; \tag{16}$$

значение ψ, соответствующее главному углу падения, обозначим через $\bar{\psi}$.

Используя (1) и (10) и зная постоянные металла n и κ, легко найти зависимость величины $P (= \operatorname{tg}\psi)$ и Δ от θ_i. На рис. 13 2, *а* она показана для типичного случая. На рис. 13.2, *б* для сравнения приведены аналогичные кривые,

*) Мы пишем $-i\Delta$, а не $+i\Delta$ в экспоненте правой части (13), чтобы облегчить сравнение с результатами, приведенными в § 1 5

**) Тогда α_r' равен углу ψ, введенному в (16).

относящиеся к случаю отражения от прозрачного диэлектрика. Для света, отраженного от поверхности металла, отсутствует тот резкий скачок Δ от $-\pi$ до 0, который происходит при отражении света под углом Брюстера от прозрачного диэлектрика. Отсутствует также острый максимум $\operatorname{tg}\psi$, равный бесконечности, и кривая для металла оказывается довольно гладкой и имеет сравнительно широкий максимум. Угол падения, при котором достигается этот максимум, иногда называют *углом наибольшей поляризации*. Он очень близок к главному углу падения $\bar{\theta}_i$. Обычно допускают, что указанный максимум находится точно при $\bar{\theta}_i$. Это предположение почти всегда оправдывается, если $n^2(1+\kappa^2)\gg 1$, что обычно и наблюдается (см. ниже табл. 13.2). Однако в общем случае эти два угла различны. Например, для серебра при длине волны 3280 Å величина $n^2(1+\kappa^2)$ мала; $\bar{\theta}_i = 47{,}8^\circ$ и $\psi = 58{,}2^\circ$, тогда как максимум $\operatorname{tg}\psi$ приблизительно соответствует $\theta_i = 40^\circ$ и $\psi_{\text{макс}} = 60{,}5^\circ$.

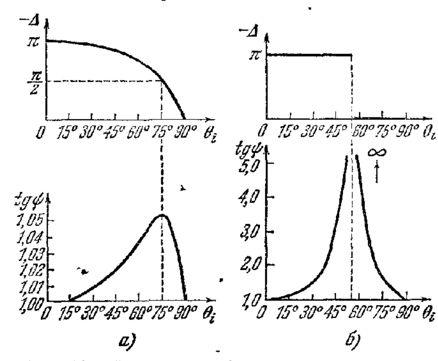


Рис. 13.2. Величины $-\Delta=\varphi_\perp-\varphi_\parallel$ и $P=\operatorname{tg}\psi=\rho_\perp/\rho_\parallel$, характеризующие изменение состояния поляризации света при отражении от типичного металла (*а*) и прозрачного диэлектрика (*б*).

Вообще говоря, проблема заключается не в том, чтобы найти ψ и Δ по известным значениям n и κ, а чтобы определить n и κ из экспериментально наблюдаемых амплитуды и фазы света, отраженного от металла.

Так как все величины $R_\parallel$, $R_\perp$, $\varphi_\parallel$, $\varphi_\perp$, ψ и Δ являются функциями θ_i, а также n и κ, то измерение любых двух из них для какого-то значения угла падения θ_i позволит, вообще говоря, найти n и κ. Во многих экспериментах определяют последние две из этого ряда величин, и поэтому мы выведем фундаментальные зависимости n и κ от ψ и Δ. Из (1) и (13) имеем

$$\frac{1-P\exp(-i\Delta)}{1+P\exp(-i\Delta)} = -\frac{\cos\theta_i\cos\theta_t}{\sin\theta_i\sin\theta_t} = -\frac{\sqrt{\hat{n}^2-\sin^2\theta_i}}{\sin\theta_i\operatorname{tg}\theta_i}. \tag{17}$$

Так как $P=\operatorname{tg}\psi$, левую часть уравнения (17) можно привести к виду

$$\frac{1-P\exp(-i\Delta)}{1+P\exp(-i\Delta)} = \frac{1-\exp(-i\Delta)\operatorname{tg}\psi}{1+\exp(-i\Delta)\operatorname{tg}\psi} = \frac{\cos 2\psi + i\sin 2\psi\sin\Delta}{1+\sin 2\psi\cos\Delta}. \tag{18}$$

Из (17) и (18) имеем

$$\frac{\sqrt{\hat{n}^2-\sin^2\theta_i}}{\sin\theta_i\operatorname{tg}\theta_i} = -\frac{\cos 2\psi + i\sin 2\psi\sin\Delta}{1+\sin 2\psi\cos\Delta}. \tag{19}$$

В видимой области спектра обычно

$$n^2(1+\kappa^2)\gg 1, \tag{20}$$

и поэтому величиной $\sin^2\theta_i$ можно пренебречь по сравнению с $\hat{n}^2$. Тогда

$$\frac{\hat{n}}{\sin\theta_i\operatorname{tg}\theta_i} = \frac{n(1+i\kappa)}{\sin\theta_i\operatorname{tg}\theta_i} \approx -\frac{\cos 2\psi + i\sin 2\psi\sin\Delta}{1+\sin 2\psi\cos\Delta}. \tag{21}$$

Приравнивая вещественные части, найдем

$$n \approx -\frac{\sin\theta_i\operatorname{tg}\theta_i\cos 2\psi}{1+\sin 2\psi\cos\Delta}. \tag{22а}$$

Приравнивая мнимые части и используя (22а), получим

$$\kappa \approx \operatorname{tg} 2\psi\sin\Delta. \tag{22б}$$

Эти выражения дают возможность рассчитать оптические постоянные n и κ по измеренным значениям ψ и Δ при любом угле падения. В частном случае,

если измерения проводятся при главном угле падения $\bar{\theta}_i$, то $\Delta=-\pi/2$, $\psi=\bar{\psi}$ и уравнения (22а) и (22б) переходят в

$$n\approx-\sin\bar{\theta}_i\,\mathrm{tg}\,\bar{\theta}_i\cos 2\bar{\psi}, \tag{23а}$$

$$\kappa\approx-\mathrm{tg}\,2\bar{\psi}. \tag{23б}$$

Иногда оказываются полезными и другие соотношения для n и κ. Без использования условия (20) найдем, возводя в квадрат (19),

$$\frac{\hat{n}^2-\sin^2\theta_i}{\sin^2\theta_i\,\mathrm{tg}^2\theta_i}=\frac{\cos^2 2\psi-\sin^2 2\psi\sin^2\Delta+i\sin 4\psi\sin\Delta}{(1+\sin 2\psi\cos\Delta)^2}. \tag{24}$$

Если подставить

$$\hat{n}^2=n^2(1-\kappa^2)+2in^2\kappa$$

и приравнять отдельно вещественную и мнимую части, то мы получим

$$n^2(1-\kappa^2)=\sin^2\theta_i\left\{1+\frac{\mathrm{tg}^2\theta_i(\cos^2 2\psi-\sin^2 2\psi\sin^2\Delta)}{(1+\sin 2\psi\cos\Delta)^2}\right\}, \tag{25а}$$

$$2n^2\kappa=\frac{\sin^2\theta_i\,\mathrm{tg}^2\theta_i\sin 4\psi\sin\Delta}{(1+\sin 2\psi\cos\Delta)^2}. \tag{25б}$$

В частности, при главном угле падения ($\theta_i=\bar{\theta}_i$, $\Delta=-\pi/2$) эти уравнения принимают вид *)

$$n^2(1-\kappa^2)=\sin^2\bar{\theta}_i(1+\mathrm{tg}^2\bar{\theta}_i\cos 4\bar{\psi}), \tag{26а}$$

$$2n^2\kappa=-\sin^2\bar{\theta}_i\,\mathrm{tg}^2\bar{\theta}_i\sin 4\bar{\psi}. \tag{26б}$$

Формулы (25) дают не n и κ, а комбинации $n^2(1-\kappa^2)$ и $n^2\kappa$. Обращаясь к (13.1.16), мы видим, что эти величины имеют простой физический смысл. При $\mu=1$ (что всегда справедливо в оптическом диапазоне) $n^2(1-\kappa^2)$ — диэлектрическая постоянная, а $n^2\kappa$ — отношение проводимости к частоте. Из значений этих величин и, в частности, из их дисперсии можно получить информацию о структуре металлов (см. ниже, § 13.3).

До сих пор наш анализ относился к амплитудам компонент отраженного света, но, как мы вскоре увидим, можно получить полезную информацию и из сравнения интенсивностей отраженного и падающего света, особенно при больших длинах волн. Для нормального падения ($\theta_i=0$) различие между $R_\parallel$ и $R_\perp$ исчезает, плоскость падения становится неопределенной, и можно написать

$$\mathcal{R}=\left|\frac{R_\parallel}{A_\parallel}\right|^2=\left|\frac{R_\perp}{A_\perp}\right|^2. \tag{27}$$

Если использовать (1) и (10) (или заменить n на $\hat{n}$ в (1.5.23)), то мы получим

$$\mathcal{R}=\left|\frac{\hat{n}-1}{\hat{n}+1}\right|^2=\frac{n^2(1+\kappa^2)+1-2n}{n^2(1+\kappa^2)+1+2n}. \tag{28}$$

Оптические постоянные многих металлов определялись с помощью измерений в отраженном свете. В табл. 13.2 приведены значения постоянных, найденные различными исследователями для длины волны, соответствующей желтой области видимого спектра. Металлы расположены в порядке убывания их отражательной способности $\mathcal{R}$. Отметим, что во всех случаях $n<n\kappa$, и поэтому, согласно (13.1.16а), $\mu\varepsilon$ и, следовательно, ε отрицательны (поскольку в оптическом диапазоне $\mu\approx1$). На первый взгляд представляется, что отрицательной диэлектрической проницаемости нельзя приписать физический смысл.

*) Уравнения (23) и (26), которые справедливы только при главном угле падения, проще, чем более общие выражения (22) или (25), и лишь по этой причине многие экспериментаторы ограничивались измерениями при главном угле. Для других углов падения экспериментальная точность может быть выше. Выбор подходящего угла падения рассматривается в работах [2—6].

Позже мы покажем, что это не так и что отрицательное значение ε можно объяснить, исходя из некоторых простых предположений об электронном механизме проводимости. Из табл. 13.2 может показаться, что значения $n<1$ всегда связаны с высокой отражательной способностью, однако это не является общим правилом *).

Таблица 13.2

Оптические постоянные металлов для длины волны $\lambda = 5893$ Å (D-линия натрия) [7]

Металл	n	$n\kappa$	$\mathcal{R}$	Исследователь	
Натрий твердый	0,044	2,42	0,97	Дункан	1913
Серебро массивное	0,20	3,44	0,94	Оппитц	1917
Магний массивный	0,37	4,42	0,93	Друде	1890
Калий расплавленный	0,084	1,81	0,92	Натансон	1928
Кадмий массивный	1,13	5,01	0,84	Друде	1890
Алюминий массивный	1,44	5,23	0,83	Друде	1890
Олово массивное	1,48	5,25	0,83	Друде	1890
Золото электролитическое	0,47	2,83	0,82	Мейер	1910
Ртуть жидкая	1,60	4,80	0,77	Ловери и Мур	1932
Цинк массивный	1,93	4,66	0,75	Мейер	1910
Медь массивная	0,62	2,57	0,73	Оппитц	1917
Галлий монокристалл	3,69	5,43	0,71	Ланге	1935
Сурьма массивная	3,04	4,94	0,70	Друде	1890
Кобальт массивный	2,12	4,04	0,68	Майнор	1904
Никель электролитический	1,58	3,42	0,66	Мейер	1910
Марганец массивный	2,41	3,88	0,64	Литтлтон	1911
Свинец массивный	2,01	3,48	0,62	Друде	1890
Платина электролитическая	2,63	3,54	0,59	Мейер	1910
Рений массивный	3,00	3,44	0,57	Ланге	1935
Вольфрам массивный	3,46	3,25	0,54	Литтлтон	1912
Висмут массивный	1,78	2,80	0,54	Мейер	1910
Железо испаренное	1,51	1,63	0,33	Мейер	1910

Значения, указанные в табл. 13.2, не согласуются с формулами (13.1.16б) или (13.1.24). Например, для меди $\sigma = 5{,}14\cdot 10^{17}$ *сек*$^{-1}$, так что для света с длиной волны 5893 Å ($\nu \sim 5\cdot 10^{14}$ *сек*$^{-1}$) $\sigma/\nu \sim 10^3$, тогда как, согласно таблице, $n^2\kappa = 1{,}57$. Кроме того, изучение зависимости оптических постоянных от частоты показывает значительно более сложное поведение, чем предсказанное нашей формулой (см. ниже, рис. 13.3). Таким образом, необходимо сделать заключение, что наша теория не адекватна, когда она применяется к излучению в видимой области электромагнитного спектра. Это расхождение между теорией и экспериментом, по-видимому, не так удивительно, если вспомнить, что даже для прозрачных сред соотношение, связывающее материальные постоянные с показателем преломления (соотношение Максвелла $\mu\varepsilon = n^2$), имеет ограниченную применимость. Объяснение аналогично данному ранее: мы не находим подтверждения предположению, что ε, μ и σ являются действительно постоянными и должны рассматривать их как функции частоты; следовательно, и показатель преломления, и показатель поглощения также будут зависеть от частоты. Единственное различие в механизме дисперсии заключается в том, что в прозрачной среде дисперсия связана с вынужденными колебаниями связанных электронов, тогда как в металле она связана с вынужденными колебаниями свободных электронов. Мы подробно обсудим это в § 13.3; здесь мы отметим лишь, что если интерпретировать ε как статическую диэлектрическую проницаемость и σ — как статическую проводимость, то можно ожидать, что

*) Когда $n<1$, вещественная фазовая скорость c/n превышает скорость света в вакууме, но, как показано в конце п. 1.3.3, это не противоречит теории относительности.

наша теория окажется справедливой только для достаточно медленных колебаний (длинных волн). Хаген и Рубенс [8] показали, что для инфракрасного излучения с длиной волны $\lambda > 1{,}2\cdot 10^{-3}$ *см* отражательная способность, рассчитанная по данным о статической проводимости, находится в хорошем согласии с экспериментом.

Если подставить n и $n\kappa$ из (13.1.24) в уравнение (28), оно примет вид (считается, что $\mu = 1$)

$$\mathcal{R} = \frac{2\dfrac{\sigma}{\nu}+1-2\sqrt{\sigma/\nu}}{2\dfrac{\sigma}{\nu}+1+2\sqrt{\sigma/\nu}}. \tag{29}$$

Когда ν/σ достаточно мало, можно пренебречь единицей по сравнению с другими членами и разложить (29) в ряд по степеням $\sqrt{\nu/\sigma}$. Тогда мы получим

$$\mathcal{R} \approx 1 - 2\sqrt{\nu/\sigma} + \ldots \tag{30}$$

Хаген и Рубенс нашли, что при $\lambda = 1{,}2\cdot 10^{-3}$ *см* для меди $1-\mathcal{R} = 1{,}6\cdot 10^{-2}$, тогда как использование в (30) статической электрической проводимости дает $1-\mathcal{R} = 1{,}4\cdot 10^{-2}$.

При увеличении длины волны величина $\mathcal{R}$ становится почти равной единице, так что трудно измерить $1-\mathcal{R}$ с достаточной точностью. Однако Хаген и Рубенс получили полезные оценки косвенным методом. Согласно *закону Кирхгофа* для теплового излучения *) отношение излучательной способности E_ν к поглощательной способности A_ν какого-либо тела **) зависит лишь от частоты излучения ν и температуры T этого тела, но не от его природы, т. е.

$$E_\nu/A_\nu = K(\nu, T). \tag{31}$$

Здесь $K(\nu, T)$ — универсальная функция ν и T. Очевидно, K равно излучательной способности тела, поглощательная способность которого равна единице (так называемое *абсолютно черное тело*). Рассмотрим металлический образец такой толщины, что вся энергия падающего излучения, которая не отражается, поглощается внутри образца. Тогда

$$A_\nu = 1-\mathcal{R}, \tag{32}$$

и из (30), (31) и (32) получим

$$A_\nu = \frac{E_\nu}{K(\nu, T)} = 2\sqrt{\frac{\nu}{\sigma}}, \tag{33}$$

или

$$\sqrt{\sigma}\,E_\nu = 2\sqrt{\nu}\,K(\nu, T). \tag{34}$$

Правая часть последнего уравнения не зависит от природы металла. Это хорошо известная функция ν и T, причем функция $K(\nu, T)$ точно известна и из эксперимента, и из теории и выражается знаменитой формулой Планка ***).

Отсюда следует, что справедливость формулы (30) можно проверить даже тогда, когда отражательная способность $\mathcal{R}$ очень близка к единице. Определяя проводимость σ и излучательную способность E_ν в зависимости от частоты и температуры, легко установить, удовлетворяет ли произведение $\sqrt{\sigma}E_\nu$ уравнению (34). Хаген и Рубенс, используя так называемые *остаточные лучи*, подтвердили справедливость уравнения (34) для длинного инфракрасного излучения. Остаточные лучи — это лучи, которые остаются из всего широкого

*) См., например, [9] или [10].

**) Под излучательной способностью понимается лучистая энергия, испускаемая телом в единицу времени, под поглощательной способностью — часть лучистой энергии, падающей на тело, которая поглощается им.

***) См., например, [11].

спектрального интервала после ряда отражений от определенных кристаллов, например флуорита, каменной соли или сильвина. Указанные вещества обладают резко выраженными максимумами поглощения в спектральной области от $\lambda=22{,}9$ *мк* до 63 *мк* и, следовательно, (см. (28)) высоким селективным отражением для таких длин волн.

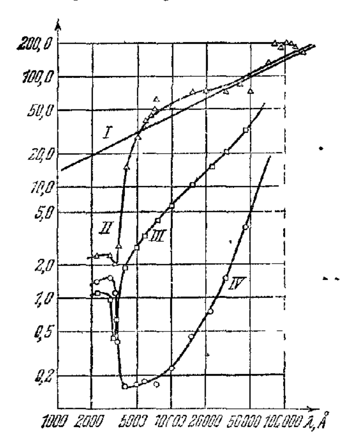


Рис. 13.3. Зависимость оптических постоянных серебра от длины волны. *I* — теоретическая кривая *nк*, *II* — экспериментальная кривая $2/(1-\mathcal{R})$, *III* — экспериментальная кривая *nк*, *IV* — экспериментальная кривая *n*. Логарифмический масштаб

На рис. 13.3 приведены экспериментальные кривые, иллюстрирующие зависимость *n* и *nк* от длины волны для серебра. Для сравнения показана также теоретическая кривая, рассчитанная по формуле (13.1 24). Здесь принят логарифмический масштаб, и поэтому теоретический график имеет вид прямой

$$\lg n \approx \lg n\kappa \approx {}^1/_2 \lg \lambda + C,$$

где $C=\lg\sqrt{\mu\sigma/c}$. Воспользовавшись (13.1.24) и (30), можно также выразить *n* и *nк* через отражательную способность (для длинных волн) в виде

$$n \approx n\kappa \approx \frac{2}{1-\mathcal{R}}. \qquad (35)$$

График функции $2/(1-\mathcal{R})$ также приведен для сравнения на рисунке. Мы видим, что экспериментальная кривая для *nк* имеет резкий минимум вблизи $\lambda=3000$ Å, а кривая для *n* — значительно более плоский минимум вблизи $\lambda=5000$ Å В области λ - 3300 Å отражательная способность серебра очень мала.

При увеличении длины волны экспериментальные кривые приближаются к теоретической кривой, рассчитанной по измеренной электропроводности.

§ 13.3. Элементарная электронная теория оптических постоянных металлов

Формальная теория, развитая в предыдущих разделах, рассматривает металл как непрерывную проводящую среду, характеризуемую диэлектрической проницаемостью ε, магнитной проницаемостью μ и электрической проводимостью σ. Как мы видели, эту теорию нельзя считать достаточно удовлетворительной в оптической области электромагнитного спектра. Лучшей представляется теория, которая рассматривает металл не как бесструктурную среду, а как систему фиксированных ионизированных атомов металла и свободно (или почти свободно) движущихся между ними электронов. В соответствии с (13 1.7) *полный* свободный заряд в металле равен нулю, но можно предположить, что именно подвижность свободных электронов служит причиной зависимости оптических постоянных металла от частоты падающего излучения. Если рассматривать электроны как частицы, подчиняющиеся законам классической механики, то такая теория по-прежнему не будет удовлетворительно описывать все наблюдаемые особенности поведения металла. Полное согласие с экспериментом способен дать лишь подход, основанный на квантовой механике. Однако такое рассмотрение выходит за рамки настоящей книги. Здесь мы лишь кратко укажем, как классическая атомистическая модель объясняет, по крайней мере до известной степени, те расхождения, с которыми мы столкнулись.

Рассмотрим электроны, движущиеся через решетку, образованную ионами. Они движутся практически свободно, за исключением коротких периодов столкновений с ионами, в течение которых между ионами и электронами происходит обмен небольшими количествами энергии. Мы не можем входить в детальное исследование этого процесса, который необходимо рассматривать с помощью статистических методов кинетической теории газов. Весьма правдоподобным представляется вывод о том, что указанный процесс в среднем можно характеризовать определенной силой трения, пропорциональной и противоположной по направлению скорости движения некоторой квазичастицы. Поведение этой квазичастицы, являющейся моделью электрона, отражает среднее поведение всей системы электронов. Уравнение движения такого модельного электрона в электрическом поле $\mathbf{E}$ запишется в виде

$$m\ddot{\mathbf{r}}+m\beta\dot{\mathbf{r}}=e\mathbf{E}, \tag{1}$$

где m — масса, e — заряд электрона и β — показатель затухания, отнесенный к единице массы. Для того чтобы понять смысл постоянной β, рассмотрим вначале случай, когда электрическое поле отсутствует. Если $\mathbf{E}=0$, то

$$\ddot{\mathbf{r}}+\beta\dot{\mathbf{r}}=0. \tag{2}$$

Решение этого уравнения имеет вид

$$\mathbf{r}=\mathbf{r}_0-\frac{1}{\beta}\mathbf{v}_0\exp(-\beta t), \qquad \dot{\mathbf{r}}=\mathbf{v}=\mathbf{v}_0\exp(-\beta t). \tag{3}$$

Мы видим, что в данном случае электрон экспоненциально замедляется, начиная от скорости $\mathbf{v}_0$, причем показатель затухания равен β. Время $\tau=1/\beta$ называется *временем затухания*.

Рассмотрим теперь гармоническое во времени поле $\mathbf{E}=\mathbf{E}_0\exp(-i\omega t)$. Тогда решение уравнения (1) является суммой двух членов, один из которых представляет затухающее движение (решение однородного уравнения (2)), а другой описывает периодическое движение

$$\mathbf{r}=-\frac{e}{m(\omega^2+i\beta\omega)}\mathbf{E}. \tag{4}$$

Периодическое движение вносит в поляризацию среды вклад $\mathbf{p}=e\mathbf{r}$. Если в единице объема содержится N свободных электронов, то полная поляризация $\mathbf{P}$, обусловленная всеми электронами, равна

$$\mathbf{P}=N\mathbf{p}=-\frac{Ne^2}{m(\omega^2+i\beta\omega)}\mathbf{E}. \tag{5}$$

Так как $\hat{\varepsilon}\mathbf{E}=\mathbf{E}+4\pi\mathbf{P}$, отсюда следует, что

$$\hat{\varepsilon}=1-\frac{4\pi Ne^2}{m(\omega^2+i\beta\omega)}. \tag{6}$$

Поскольку (при $\mu=1$) $\hat{\varepsilon}=n^2(1-\kappa^2)+2in^2\kappa$, то, приравнивая отдельно вещественные и мнимые части и используя уравнения (13.1.16), получим

$$\varepsilon=n^2(1-\kappa^2)=1-\frac{4\pi Ne^2}{m(\omega^2+\beta^2)}, \tag{7а}$$

$$\frac{\sigma}{\nu}=n^2\kappa=\frac{2\pi Ne^2\beta}{m\omega(\omega^2+\beta^2)}. \tag{7б}$$

Если β невелико, то для достаточно малых значений ω величина ε отрицательна. Критическое значение ω_c, при котором ε меняет знак, определяется выражением

$$\omega_c^2=\frac{4\pi Ne^2}{m}-\beta^2. \tag{8}$$

Используя это выражение, можно переписать уравнения (7) в виде

$$\varepsilon=n^2(1-\kappa^2)=1-\frac{\omega_c^2+\beta^2}{\omega^2+\beta^2}, \tag{9а}$$

$$\frac{\sigma}{\nu}=n^2\kappa=\frac{\beta(\omega_c^2+\beta^2)}{2\omega(\omega^2+\beta^2)}. \tag{9б}$$

Предположим, что ω_c^2 значительно больше β^2; тогда вместо (8) мы можем написать

$$\omega_c^2 \approx \frac{4\pi N e^2}{m}. \quad (10)$$

Если ограничиться также достаточно высокими частотами ($\omega^2 \gg \beta^2$), то вместо (9) получим более простые формулы

$$\varepsilon = n^2(1-\kappa^2) \approx 1 - \left(\frac{\omega_c}{\omega}\right)^2, \quad (11а)$$

$$\frac{\sigma}{\nu} = n^2\kappa \approx \frac{\beta}{2\omega}\left(\frac{\omega_c}{\omega}\right)^2. \quad (11б)$$

Из (11а) следует, что если $\omega < \omega_c$, то ε отрицательно и $\kappa > 1$. Мы видим теперь, что отрицательное значение вещественной части диэлектрической проницаемости имеет вполне определенный смысл. Оно указывает, что поле, возникающее в результате колебаний электронов, находится в противофазе с возбуждающим полем. Отражательная способность, определяемая выражением (13.2 28), в рассматриваемом случае очень велика. Если же $\omega > \omega_c$, то ε принимает положительное значение, κ не превышает единицы и отражательная способность падает до чрезвычайно низкого значения. Среда становится довольно прозрачной, похожей на диэлектрик.

Именно такие явления наблюдаются в случае щелочных металлов. В области длинных волн они непрозрачны и обладают высоким отражением, тогда как при некоторой критической длине волны в видимой или ультрафиолетовой области спектра они становятся прозрачными и сравнительно мало поглощают. Во второй строке табл. 13.3 указаны экспериментально найденные длины

Таблица 13.3

Критические длины волн λ_c

Металл	Литий	Натрий	Калий	Рубидий	Цезий
$(\lambda_c)_{набл}$, Å	2050	2100	3150	3600	4400
$(\lambda_c)_{расч}$, Å	1500	2100	2900	3200	3600
$\frac{N_{эфф}}{N}$	0,54	1,00	0,85	0,79	0,67

волн, при которых происходит такой переход. В третьей строке приведены критические длины волн $\lambda_c = 2\pi c/\omega_c$, рассчитанные по приближенной формуле (10), где число свободных электронов считается равным числу N атомов в единице объема. Как мы видим, для всех металлов, кроме натрия, наблюдаемые и рассчитанные значения критических длин волн различны. Последняя строчка содержит отношение концентрации «эффективных» электронов $N_{эфф}$ к концентрации атомов, найденное из соотношения

$$\frac{N_{эфф}}{N} = \frac{(\lambda_c)^2_{расч}}{(\lambda_c)^2_{набл}}. \quad (12)$$

Как мы видим, это число порядка единицы, хотя и заметно меньше (для всех металлов, кроме натрия). Итак, элементарная теория дает правильный порядок величины параметров, но не описывает деталей явления.

Предложенную выше теорию можно улучшить, если вместо грубых приближений (11) использовать более точные формулы (9), содержащие показатель затухания β. Однако расчет β невозможен в рамках классической теории и требует привлечения квантовой механики. Еще одним указанием на неудовлетворительность классической модели служит появление другой области

непрозрачности при коротких длинах волн, которая не предсказывается нашими формулами. Наличие этой области можно объяснить на основе квантовой механики как результат поглощения, связанного с внутренним фотоэффектом. За счет энергии падающего света электроны с их нормальных орбит переходят в возбужденные состояния. Однако такие соображения выходят за рамки настоящей книги.

§ 13.4. Распространение волн в слоистой проводящей среде, Теория металлических пленок

В § 1.6 мы изучали распространение электромагнитных волн в слоистых диэлектрических средах, т. е. в диэлектрических средах, оптические свойства которых зависят лишь от одной из декартовых координат. Сейчас мы кратко рассмотрим обобщение теории на слоистые среды, содержащие поглощающие элементы. Таким образом, мы предположим, что в дополнение к ε и μ, зависящим лишь от одной координаты, имеется конечная проводимость σ, которая также является функцией только этой координаты.

В начале § 13.2 мы уже указывали, что все формулы гл. 1, которые содержат лишь линейные соотношения между компонентами векторов поля гармонической волны, остаются справедливыми и для проводящих сред, если вещественную диэлектрическую проницаемость ε и вещественное волновое число k заменить соответственно на комплексную диэлектрическую проницаемость $\hat{\varepsilon}=\varepsilon+i4\pi\sigma/\omega$ и комплексное волновое число $\hat{k}=\omega\sqrt{\mu(\varepsilon+i4\pi\sigma/\omega)}/c$. Следовательно, мы можем использовать основные формулы теории слоистых диэлектрических сред, выведенные в § 1.6, при условии, что в соответствующих формулах будет сделано указанное формальное изменение. Отсюда, в частности, следует, что слоистую поглощающую среду можно характеризовать матрицей 2×2. Однако элементы этой матрицы, в отличие от матрицы, относящейся к диэлектрической слоистой среде, уже не являются вещественными или чисто мнимыми, а представляют собой комплексные числа, содержащие как вещественную, так и мнимую части.

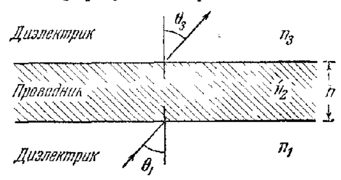


Рис. 13.4. Поглощающая пленка, расположенная между двумя диэлектрическими средами.

Для иллюстрации теории подробно рассмотрим два случая, представляющих практический интерес.

13.4.1. Поглощающая пленка на прозрачной подложке. Рассмотрим плоскопараллельную поглощающую пленку, расположенную между двумя диэлектрическими средами (рис. 13.4). Формулы, относящиеся к отражению и пропусканию пленкой плоской монохроматической волны, получаются из уравнений (1.6.55) — (1.6.58) при замене n_2 на $\hat{n}_2=n_2(1+i\kappa_2)$. Удобно положить

$$\hat{n}_2\cos\theta_2=u_2+iv_2, \tag{1}$$

где u_2 и v_2 — вещественные величины. Легко выразить u_2 и v_2 через угол падения и постоянные, характеризующие оптические свойства первой и второй сред. Возводя в квадрат (1) и используя закон преломления $\hat{n}_2\sin\theta_2=n_1\sin\theta_1$, получим

$$(u_2+iv_2)^2=\hat{n}_2^2-n_1^2\sin^2\theta_1. \tag{2}$$

Приравнивая отдельно вещественную и мнимую части, найдем

$$\left.\begin{aligned} u_2^2-v_2^2&=n_2^2(1-\kappa_2^2)-n_1^2\sin^2\theta_1,\\ u_2v_2&=n_2^2\kappa_2. \end{aligned}\right\} \tag{3}$$

Отсюда следует, что

$$\left.\begin{aligned}2u_2^2&=n_2^2(1-\kappa_2^2)-n_1^2\sin^2\theta_1+\sqrt{[n_2^2(1-\kappa_2^2)-n_1^2\sin^2\theta_1]^2+4n_2^4\kappa_2^2},\\ 2v_2^2&=-[n_2^2(1-\kappa_2^2)-n_1^2\sin^2\theta_1]+\sqrt{[n_2^2(1-\kappa_2^2)-n_1^2\sin^2\theta_1]^2+4n_2^4\kappa_2^2}.\end{aligned}\right\}\tag{4}$$

Далее необходимо найти коэффициенты отражения и пропускания для поверхностей раздела 1—2 и 2—3 соответственно, поскольку эти коэффициенты входят в формулы для коэффициентов отражения и пропускания пленки. Мы отдельно рассмотрим случаи, когда электрический вектор падающей волны перпендикулярен и параллелен плоскости падения.

Электрический вектор перпендикулярен плоскости падения (TE волна).

В этом случае при замене в (1.6.55) величины $n_2\cos\theta_2$ на $\hat{n}_2\cos\theta_2=u_2+iv_2$ получим

$$r_{12}=\rho_{12}\exp(i\varphi_{12})=\frac{n_1\cos\theta_1-(u_2+iv_2)}{n_1\cos\theta_1+(u_2+iv_2)}.\tag{5}$$

Позже нам понадобятся выражения для амплитуды ρ_{12} и изменения фазы φ_{12} в явном виде. Из (5) найдем

$$\rho_{12}^2=\frac{(n_1\cos\theta_1-u_2)^2+v_2^2}{(n_1\cos\theta_1+u_2)^2+v_2^2},\qquad \operatorname{tg}\varphi_{12}=\frac{2v_2n_1\cos\theta_1}{u_2^2+v_2^2-n_1^2\cos^2\theta_1}.\tag{6}$$

Пропускание на первой поверхности раздела, согласно (1.6.56), запишется в виде

$$t_{12}=\tau_{12}\exp(i\chi_{12})=\frac{2n_1\cos\theta_1}{n_1\cos\theta_1+u_2+iv_2}.\tag{7}$$

Это дает

$$\tau_{12}^2=\frac{(2n_1\cos\theta_1)^2}{(n_1\cos\theta_1+u_2)^2+v_2^2},\qquad \operatorname{tg}\chi_{12}=-\frac{v_2}{n_1\cos\theta_1+u_2}.\tag{8}$$

Совершенно аналогичным образом мы получим следующие выражения для отражения и пропускания на второй поверхности раздела:

$$\rho_{23}^2=\frac{(n_3\cos\theta_3-u_2)^2+v_2^2}{(n_3\cos\theta_3+u_2)^2+v_2^2},\qquad \operatorname{tg}\varphi_{23}=\frac{2v_2n_3\cos\theta_3}{u_2^2+v_2^2-n_3^2\cos^2\theta_3},\tag{9}$$

$$\tau_{23}^2=\frac{4(u_2^2+v_2^2)}{(n_3\cos\theta_3+u_2)^2+v_2^2},\qquad \operatorname{tg}\chi_{23}=\frac{v_2n_3\cos\theta_3}{u_2^2+v_2^2+u_2n_3\cos\theta_3}.\tag{10}$$

На основании закона преломления $n_1\sin\theta_1=\hat{n}_2\sin\theta_2$, $\hat{n}_2\sin\theta_2=n_3\sin\theta_3$; поэтому угол θ_3 определяется через θ_1 посредством соотношения

$$n_3\sin\theta_3=n_1\sin\theta_1.\tag{11}$$

Электрический вектор параллелен плоскости падения (TM волна).

В п. 1.6.3 было показано, что формулы для коэффициентов отражения и пропускания TM волны легко получить из формул для TE волны, заменяя величины $p_j=n_j\cos\theta_j$ на $q_j=\cos\theta_j/n_j$ (при этом среда считается немагнитной). Величины r и t теперь представляют отношения магнитных, а не электрических векторов. В частности, из (1.6.55) следует, что

$$r_{12}=\rho_{12}\exp(i\varphi_{12})=\frac{\frac{1}{n_1}\cos\theta_1-\frac{1}{\hat{n}_2}\cos\theta_2}{\frac{1}{n_1}\cos\theta_1+\frac{1}{\hat{n}_2}\cos\theta_2}=\frac{\hat{n}_2^2\cos\theta_1-n_1\hat{n}_2\cos\theta_2}{\hat{n}_2^2\cos\theta_1+n_1\hat{n}_2\cos\theta_2}=$$

$$=\frac{[n_2^2(1-\kappa_2^2)+2in_2^2\kappa_2]\cos\theta_1-n_1(u_2+iv_2)}{[n_2^2(1-\kappa_2^2)+2in_2^2\kappa_2]\cos\theta_1+n_1(u_2+iv_2)}.\tag{12}$$

Отсюда легко получить

$$\left.\begin{aligned}\rho_{12}^2&=\frac{[n_2^2(1-\kappa_2^2)\cos\theta_1-n_1u_2]^2+[2n_2^2\kappa_2\cos\theta_1-n_1v_2]^2}{[n_2^2(1-\kappa_2^2)\cos\theta_1+n_1u_2]^2+[2n_2^2\kappa_2\cos\theta_1+n_1v_2]^2},\\ \operatorname{tg}\varphi_{12}&=2n_1n_2^2\cos\theta_1\frac{2\kappa_2u_2-(1-\kappa_2^2)v_2}{n_2^4(1+\kappa_2^2)^2\cos^2\theta_1-n_1^2(u_2^2+v_2^2)}.\end{aligned}\right\}\tag{13}$$

Отношение t_{12} мы получим из (1.6.56), заменяя $n_j \cos\theta_j$ на $\cos\theta_j/n_j$:

$$t_{12}=\tau_{12}\exp(i\chi_{12})=\frac{\frac{2}{n_1}\cos\theta_1}{\frac{1}{n_1}\cos\theta_1+\frac{1}{\hat{n}_2}\cos\theta_2}=\frac{2\left[n_2^2(1-\kappa_2^2)+2in_2^2\kappa_2\right]\cos\theta_1}{\left[n_2^2(1-\kappa_2^2)+2in_2^2\kappa_2\right]\cos\theta_1+n_1(u_2+iv_2)}. \tag{14}$$

Отсюда

$$\left.\begin{aligned}\tau_{12}^2&=\frac{4n_2^4(1+\kappa_2^2)^2\cos^2\theta_1}{\left[n_2^2(1-\kappa_2^2)\cos\theta_1+n_1u_2\right]^2+\left[2n_2^2\kappa_2\cos\theta_1+n_1v_2\right]^2},\\ \operatorname{tg}\chi_{12}&=\frac{n_1\left[2\kappa_2u_2-(1-\kappa_2^2)v_2\right]}{n_2^2(1+\kappa_2^2)^2\cos\theta_1+n_1\left[(1-\kappa_2^2)u_2+2\kappa_2v_2\right]}.\end{aligned}\right\} \tag{15}$$

Подобным же образом мы получим следующие формулы для коэффициентов отражения и пропускания, относящихся ко второй поверхности раздела:

$$\left.\begin{aligned}\rho_{23}^2&=\frac{\left[n_2^2(1-\kappa_2^2)\cos\theta_3-n_3u_2\right]^2+\left[2n_2^2\kappa_2\cos\theta_3-n_3v_2\right]^2}{\left[n_2^2(1-\kappa_2^2)\cos\theta_3+n_3u_2\right]^2+\left[2n_2^2\kappa_2\cos\theta_3+n_3v_2\right]^2},\\ \operatorname{tg}\varphi_{23}&=2n_3n_2^2\cos\theta_3\,\frac{2\kappa_2u_2-(1-\kappa_2^2)v_2}{n_2^4(1+\kappa_2^2)^2\cos^2\theta_3-n_3^2(u_2^2+v_2^2)}\end{aligned}\right\} \tag{16}$$

и

$$\left.\begin{aligned}\tau_{23}^2&=\frac{4n_3^2(u_2^2+v_2^2)}{\left[n_3u_2+n_2^2(1-\kappa_2^2)\cos\theta_3\right]^2+\left[n_3v_2+2n_2^2\kappa_2\cos\theta_3\right]^2},\\ \operatorname{tg}\chi_{23}&=\frac{n_2^2\left[(1-\kappa_2)^2v_2-2\kappa_2u_2\right]\cos\theta_3}{n_3\left[u_2^2+v_2^2\right]+n_2^2\left[(1-\kappa_2)^2u_2+2\kappa_2v_2\right]\cos\theta_3}.\end{aligned}\right\} \tag{17}$$

Зная величины ρ_{12}, φ_{12} и т. д., сразу же можно найти комплексные коэффициенты отражения и пропускания пленки. Полезно положить

$$\eta=\frac{2\pi}{\lambda_0}h, \tag{18}$$

так что

$$\beta=\frac{2\pi}{\lambda_0}\hat{n}_2h\cos\theta_2=(u_2+iv_2)\,\eta. \tag{19}$$

Уравнения (1.6.57) и (1.6.58) теперь примут вид

$$r=\rho\exp(i\delta_r)=\frac{\rho_{12}\exp(i\varphi_{12})+\rho_{23}\exp(-2v_2\eta)\exp\left[i(\varphi_{23}+2u_2\eta)\right]}{1+\rho_{12}\rho_{23}\exp(-2v_2\eta)\exp\left[i(\varphi_{12}+\varphi_{23}+2u_2\eta)\right]}, \tag{20}$$

$$t=\tau\exp(i\delta_t)=\frac{\tau_{12}\tau_{23}\exp(-2v_2\eta)\exp\left[i(\chi_{12}+\chi_{23}+u_2\eta)\right]}{1+\rho_{12}\rho_{23}\exp(-2v_2\eta)\exp\left[i(\varphi_{12}+\varphi_{23}+2u_2\eta)\right]}. \tag{21}$$

Выражения для отражательной способности $\mathcal{R}$ и изменения фазы при отражении δ_r получаются непосредственно из (20); имеем

$$\mathcal{R}=|r|^2=\frac{\rho_{12}^2\exp(2v_2\eta)+\rho_{23}^2\exp(-2v_2\eta)+2\rho_{12}\rho_{23}\cos\left[\varphi_{23}-\varphi_{12}+2u_2\eta\right]}{\exp(2v_2\eta)+\rho_{12}^2\rho_{23}^2\exp(-2v_2\eta)+2\rho_{12}\rho_{23}\cos\left[\varphi_{12}+\varphi_{23}+2u_2\eta\right]}, \tag{22}$$

$$\operatorname{tg}\delta_r=\frac{\rho_{23}(1-\rho_{12}^2)\sin(2u_2\eta+\varphi_{23})+\rho_{12}\left[\exp(2v_2\eta)-\rho_{23}^2\exp(-2v_2\eta)\right]\sin\varphi_{12}}{\rho_{23}(1+\rho_{12}^2)\cos(2u_2\eta+\varphi_{23})+\rho_{12}\left[\exp(2v_2\eta)+\rho_{23}^2\exp(-2v_2\eta)\right]\cos\varphi_{12}}. \tag{23}$$

Эти формулы справедливы как для ТЕ волны, так и для ТМ волны. В первом случае вместо ρ и φ необходимо подставить значения, определяемые соотношениями (6) и (9), во втором — формулами (13) и (16).

Подобным же образом из (21) получаются следующие выражения для пропускательной способности $\mathcal{T}$ и изменения фазы при пропускании δ_t:

$$\mathcal{T} = \frac{n_3 \cos\theta_3}{n_1 \cos\theta_1} |t|^2 =$$

$$= \frac{n_3 \cos\theta_3}{n_1 \cos\theta_1} \frac{\tau_{12}^2 \tau_{23}^2 \exp(-2v_2\eta)}{\exp(2v_2\eta) + \rho_{12}^2 \rho_{23}^2 \exp(-2v_2\eta) + 2\rho_{12}\rho_{23}\cos[\varphi_{12} + \varphi_{23} + 2u_2\eta]}, \quad (24)$$

$$\operatorname{tg}[\delta_t - \chi_{12} - \chi_{23} + u_2\eta] = \frac{\exp(2v_2\eta)\sin 2u_2\eta - \rho_{12}\rho_{23}\sin(\varphi_{12} + \varphi_{23})}{\exp(2v_2\eta)\cos 2u_2\eta + \rho_{12}\rho_{23}\cos(\varphi_{12} + \varphi_{23})}. \quad (25)$$

Для ТМ волны множитель $(n_3 \cos\theta_3)/(n_1 \cos\theta_1)$ необходимо заменить на $(\cos\theta_3/n_3)/(\cos\theta_1/n_1)$. В эти формулы нужно подставить значения, определяемые соотношениями (6), (8) — (10) для ТЕ волны и (13), (15) — (17) для ТМ волны.

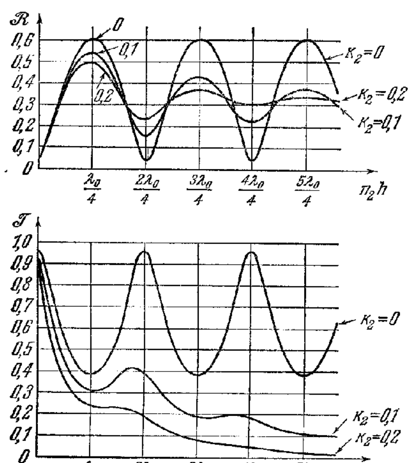


Рис. 13.5. Зависимость отражательной способности $\mathcal{R}$ и пропускательной способности $\mathcal{T}$ металлической пленки от ее оптической толщины [12].

$n_1=1,\ n_2=3{,}5,\ n_3=1{,}5\quad \kappa_1=\kappa_3=0,\ \theta_1=0.$

Напомним, что изменение фазы при отражении (δ_r) относится к первой границе, а изменение фазы при пропускании (δ_t) — ко второй.

Уравнения (22) — (25) позволяют рассчитать четыре основные величины, которые характеризуют отражение и пропускание поглощающей пленки заданной толщины с известными оптическими свойствами. На рис. 13.5 для некоторых типичных случаев показана зависимость отражения и пропускания от толщины пленки.

Для непоглощающей пленки $\mathcal{R}$ и $\mathcal{T}$ — периодические функции толщины пленки h с периодом, равным одной длине волны. Как мы видим, поглощение уменьшает величину последующих максимумов и вызывает смещение максимумов в направлении меньших толщин. В оптическом диапазоне поглощение металлов столь велико, что толщина, при которой еще имеется заметное пропускание, значительно меньше четверти длины волны *) (см. табл. 13.1). Поэтому в проходящем свете максимумы и минимумы не наблюдаются.

В оптике металлические пленки применяются главным образом для получения высокой отражательной способности, например в интерферометрах Фабри — Перо (см. п. 7.6.2). Такие пленки обычно изготовляли путем химического осаждения, но в последнее время этот метод был вытеснен методом испарения в высоком вакууме (см., например, [15] или [16]).

Наконец, рассмотрим кратко отражение и пропускание «толстой» пленки. Если толщина h и, следовательно, параметр η достаточно велики, то в (22) — (25) можно пренебречь всеми членами, которые не содержат множителя $\exp(2v_2\eta)$. Например, если $\exp(2v_2\eta) \geqslant 100$, то это, как правило, не приводит к ошибке, превышающей несколько процентов. Для такой пленки при

---

*) Упрощенные формулы, относящиеся к таким тонким пленкам, можно получить, разлагая числитель и знаменатель (22)— (25) в ряд по степеням толщины пленки и оставляя члены лишь первых нескольких степеней (см. [13]).

Оптические свойства тонких металлических пленок в видимой и инфракрасной областях спектра подробно рассмотрены в работах [14].

нормальном падении $4\pi h n_2 \kappa_2/\lambda \geqslant \ln 100 = 4{,}61$ или (опуская индекс 2)

$$\frac{h}{\lambda} \geqslant \frac{0{,}37}{n\kappa}. \tag{26}$$

Для серебряной пленки, например, $n\kappa \approx 3{,}67$ при $\lambda = 5780$ А и (26) дает $h \geqslant \lambda/10 \approx 5{,}8 \cdot 10^{-6}$ *см.*

Для толстой пленки из (22) и (24) получим

$$\mathcal{R} \approx \rho_{12}^2, \quad \mathcal{T} = \frac{n_3 \cos\theta_3}{n_1 \cos\theta_1} \tau_{12}^2 \tau_{23}^2 \exp(-4v_2\eta). \tag{27}$$

Таким образом, отражательная способность «толстой» пленки почти равна отражательной способности бесконечно толстой пленки, а ее пропускательная способность экспоненциально уменьшается с толщиной. Изменения фазы легко найти из (23) и (25):

$$\delta_r \approx \varphi_{12}, \quad \delta_t \approx \chi_{12} + \chi_{23} + u_2\eta. \tag{28}$$

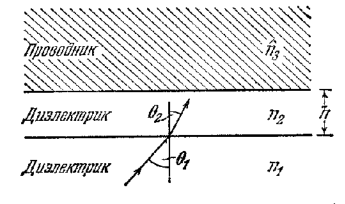


Рис. 13.6. Прозрачная пленка на поглощающей подложке.

Формулы (27) и (28) раскрывают физический смысл нашего определения «толстой» пленки, показывая, что в такой пленке можно пренебречь влиянием многолучевой интерференции.

13.4.2. Прозрачная пленка на поглощающей подложке. В качестве второго примера рассмотрим отражение от прозрачной пленки на поглощающей подложке (рис. 13.6).

В этом случае r_{12} вещественно, тогда как r_{23} комплексно. Отношение амплитуд ρ_{23} и изменение фазы φ_{23} определяются выражениями (6) или (13), где индексы 1 и 2 заменены на 2 и 3 соответственно. В соответствии с (1.6.57) найдем

$$r = \frac{r_{12} + \rho_{23} \exp[i(\varphi_{23} + 2\beta)]}{1 + r_{12}\rho_{23} \exp[i(\varphi_{23} + 2\beta)]}. \tag{29}$$

Это выражение идентично формуле (1.6.57), если в последней 2β заменено на $2\beta + \varphi_{23}$ и r_{23} на ρ_{23}. Таким образом, можно сразу же, без всякого расчета, записать выражение для отражательной способности и изменения фазы при отражении, просто выполняя указанную замену в уравнениях (1.6.59) и (1.6.61). При этом мы получим

$$\mathcal{R} = \frac{r_{12}^2 + \rho_{23}^2 + 2r_{12}\rho_{23}\cos(\varphi_{23} + 2\beta)}{1 + r_{12}^2\rho_{23}^2 + 2r_{12}\rho_{23}\cos(\varphi_{23} + 2\beta)} \tag{30}$$

и

$$\operatorname{tg}\delta_r = \frac{\rho_{23}(1 - r_{12}^2)\sin(\varphi_{23} + 2\beta)}{r_{12}(1 + \rho_{23}^2) + \rho_{23}(1 + r_{12}^2)\cos(\varphi_{23} + 2\beta)}. \tag{31}$$

Тонкие прозрачные пленки на поглощающих подложках находят многочисленные практические применения. Они используются, например, для защиты металлических зеркал от повреждения и для увеличения их отражательной способности. Эти пленки можно использовать также для уменьшения отражения от металлической поверхности. В п. 1.6.4 мы упоминали о поляризаторе, состоящем из диэлектрической пленки на диэлектрической подложке, для которого $\mathcal{R}_\parallel = 0$, а $\mathcal{R}_\perp$ достаточно велико. С металлической подложкой можно получить либо $\mathcal{R}_\parallel = 0$, либо $\mathcal{R}_\perp = 0$ [17].

§ 13.5. Дифракция на проводящей сфере. Теория Ми

Металлы обладают специфическими оптическими свойствами и тогда, когда они тонко измельчены, как в коллоидных суспензиях. Достаточно вспомнить блестящие ярко красные цвета коллоидного золота в жидкостях или

стеклах. Эти явления представляют большой интерес, так как в них проявляются вместе преломление, поглощение и дифракция.

Будь металлические частицы идеальными проводниками, мы имели бы дело с проблемой чистой дифракции. Однако мы не рассматривали этот вопрос с указанной точки зрения в главах по дифракции, поскольку для физики особый интерес представляют именно эффекты, обусловленные частичным проникновением света в частицы. В данном случае важную роль играет поглощение, и поэтому более уместно рассмотреть этот вопрос в настоящей главе. Наши выражения будут содержать в качестве предельного случая ($\kappa \rightarrow 0$) и соответствующие результаты, относящиеся к диэлектрическим сферам.

Из ранних исследователей, изучавших оптические свойства металлических частиц, необходимо упомянуть Максвелла Гарнетта [18]. Он рассматривал прохождение света через диэлектрическую среду, содержащую в объеме с линейными размерами порядка длины волны много небольших металлических сфер. С помощью формулы Лоренц — Лоренца (см. (2.3.17)) Гарнетт показал, что такая система эквивалентна среде с определенным комплексным показателем преломления $\hat{n}' = n'(1 + i\kappa')$, и выразил n' и κ' через показатели n и κ, характеризующие металлические сферы. При этом он смог объяснить некоторые наблюдаемые особенности.

В работе, опубликованной в 1908 г., Дж. Ми [19] на основе электромагнитной теории получил строгое решение для дифракции плоской монохроматической волны на однородной сфере произвольного диаметра и состава, находящейся в однородной среде. Эквивалентное решение той же проблемы было вскоре опубликовано Дебаем [20] в статье, относящейся к давлению света (т. е. механической силе, вызываемой светом) на проводящую сферу. Затем различные аспекты этой проблемы рассматривались многими авторами *).

Хотя решение, предложенное Ми, получено для дифракции на одной сфере, оно применимо также к дифракции на любом числе сфер при условии, что все они имеют одинаковый диаметр и одинаковый состав, распределены хаотически и находятся друг от друга на расстояниях, больших по сравнению с длиной волны. При таких условиях световые пучки, рассеянные сферами, не когерентны, а полная рассеянная энергия равна произведению энергии, рассеянной одной сферой, на общее число сфер. Здесь следует отметить, что решение Ми имеет большое практическое значение и его можно применить к самым разным задачам; помимо вопроса о цветах металлических суспензий, можно упомянуть такие приложения, как изучение атмосферной пыли, межзвездных частиц или коллоидов, теория радуги, солнечная корона, влияние облаков и туманов на пропускание света и т. д.

Прежде чем переходить к выводу формулы Ми, полезно кратко пояснить применяемый метод. Мы ищем решение уравнений Максвелла, описывающих поле, возникающее при падении плоской монохроматической волны на сферическую поверхность, близ которой резко меняются свойства среды. Вводится соответствующая система криволинейных (сферических) координат, и поле представляется в виде суммы двух «подполей»; электрический вектор одного из них не содержит радиальной составляющей, у другого таким же свойством обладает магнитный вектор. В сферической системе координат уравнения Максвелла вместе с граничными условиями переходят в систему обычных дифференциальных уравнений, решение которых для двух подполей ищется в виде бесконечных рядов. В п. 13.5.1 рассматривается получение этого решения, а в п. 13.5.2 — основные его следствия. Последний раздел настоящего параграфа посвящен некоторым общим результатам, относящимся к полному коли-

*) Из более поздних исследований, касавшихся основной теории, особенно следует упомянуть работы [21, 22].

честву энергии, которое рассеивается и поглощается частицей произвольной формы, причем случай сферической частицы рассматривается подробно.

13.5.1. **Математическое решение проблемы.** *а. Представление поля через потенциалы Дебая.* Рассмотрим дифракцию плоской линейно поляризованной монохроматической волны на сфере радиуса a, находящейся в однородной изотропной среде. Предположим, что среда, в которой находится сфера, является непроводящей и что как среда, так и сфера немагнитны.

Как обычно, выделим зависимость от времени в виде множителя $\exp(-i\omega t)$. Тогда электрический и магнитный векторы и вне, и внутри сферы удовлетворяют уравнениям Максвелла в форме, не зависящей от времени, т. е.

$$\operatorname{rot} \mathbf{H} = -k_1 \mathbf{E}, \tag{1a}$$

$$\operatorname{rot} \mathbf{E} = k_2 \mathbf{H}, \tag{1б}$$

где

$$k_1 = \frac{i\omega}{c}\left(\varepsilon + i\frac{4\pi\sigma}{\omega}\right), \tag{2a}$$

$$k_2 = \frac{i\omega}{c}. \tag{2б}$$

Квадрат обычного волнового числа k (вещественного вне сферы и комплексного внутри ее) равен

$$k^2 = -k_1 k_2. \tag{3}$$

Величины, которые относятся к среде, окружающей сферу, снабдим значком I, а величины, относящиеся к сфере,— значком II. Поскольку, по предположению, среда, окружающая сферу, является непроводящей, $\sigma^{(I)} = 0$.

Воспользуемся прямоугольной системой координат с началом в центре сферы. Пусть ось z совпадает с направлением распространения волны, а ось x — с направлением ее электрического вектора (рис. 13.7).

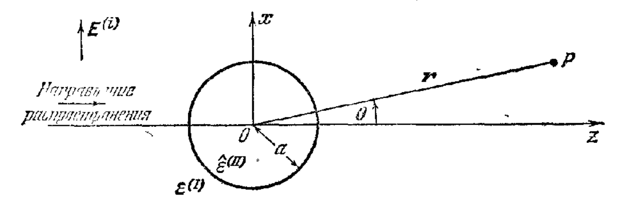


Рис. 13.7. К рассмотрению дифракции на проводящей сфере.

Если амплитуда электрического вектора падающей волны нормирована на единицу, т. е.

$$|\mathbf{E}^{(i)}| = |\exp(ik^{(I)}z)| = 1,$$

то шесть компонент векторов поля запишутся в виде

$$E_x^{(i)} = \exp(ik^{(I)}z), \quad H_y^{(i)} = \frac{ik^{(I)}}{k_2^{(I)}}\exp(ik^{(I)}z), \quad E_y^{(i)} = E_z^{(i)} = H_x^{(i)} = H_z^{(i)} = 0. \tag{4}$$

Что касается граничных условий, то в соответствии с п. 1.1.3 мы потребуем лишь, чтобы тангенциальные составляющие векторов **E** и **H** были непрерывны на поверхности сферы, т. е.

$$E_{\text{танг}}^{(I)} = E_{\text{танг}}^{(II)}, \quad H_{\text{танг}}^{(I)} = H_{\text{танг}}^{(II)}, \quad \text{когда} \quad r = a. \tag{5}$$

Тогда условие, требующее непрерывности на поверхности сферы радиальных компонент $\varepsilon\mathbf{E}$ и $\mathbf{H}$, следует из (5) и уравнений Максвелла.

Для того чтобы удовлетворить граничным условиям, необходимо предположить, что, кроме падающего поля $\mathbf{E}^{(i)}$ и $\mathbf{H}^{(i)}$ и поля внутри сферы $\mathbf{E}^{(w)}$, $\mathbf{H}^{(w)}$, имеется вторичное (рассеянное или дифрагировавшее) поле $\mathbf{E}^{(s)}$, $\mathbf{H}^{(s)}$ в среде, окружающей сферу. Таким образом, полное электрическое поле в обеих областях запишется в виде

$$\left.\begin{aligned} \mathbf{E} &= \mathbf{E}^{(i)} + \mathbf{E}^{(s)} \quad \text{вне сферы}, \\ \mathbf{E} &= \mathbf{E}^{(w)} \quad \text{внутри сферы}. \end{aligned}\right\} \tag{6}$$

Подобные же выражения справедливы и для магнитного вектора. Поля $\mathbf{E}^{(s)}$, $\mathbf{H}^{(s)}$ и $\mathbf{E}^{(w)}$, $\mathbf{H}^{(w)}$ можно считать аналогичными соответственно отраженному и проходящему полям при падении на плоскую границу (см. п. 1.5.1). Однако такая аналогия верна лишь при диаметре сферы, большом по сравнению с длиной волны. Так как граничные условия должны выполняться для любого момента времени, все шесть векторов должны обладать одинаковой зависимостью от времени [$\exp(-i\omega t)$].

Для настоящей задачи наиболее удобными криволинейными координатами являются сферические координаты r, θ и φ, определяемые выражениями

$$x = r\sin\theta\cos\varphi, \qquad y = r\sin\theta\sin\varphi, \qquad z = r\cos\theta. \tag{7}$$

При переходе от декартовой системы координат к этой новой системе компоненты произвольного вектора $\mathbf{A}$ преобразуются согласно следующим правилам (см., например, [23]):

$$\left.\begin{aligned} A_r &= A_x\sin\theta\cos\varphi + A_y\sin\theta\sin\varphi + A_z\cos\theta, \\ A_\theta &= A_x\cos\theta\cos\varphi + A_y\cos\theta\sin\varphi - A_z\sin\theta, \\ A_\varphi &= -A_x\sin\varphi + A_y\cos\varphi. \end{aligned}\right\} \tag{8}$$

Определяемые здесь компоненты не совпадают с компонентами, используемыми в абсолютном дифференциальном исчислении Риччи и Леви-Чивита. Там рассматриваются две группы различных, но эквивалентных компонент вектора $\mathbf{A}$ — контравариантные компоненты $A^{(i)}$ и ковариантные компоненты A_i. Если $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ — базисные векторы, в общем случае неортогональные и имеющие различную длину, то контравариантные компоненты (по отношению к этим базисным векторам) можно определить как коэффициенты в представлении $\mathbf{A} = A^{(1)}\mathbf{e}_1 + A^{(2)}\mathbf{e}_2 + A^{(3)}\mathbf{e}_3$, а ковариантные компоненты — как коэффициенты в представлении $\mathbf{A} = A_1\mathbf{e}^{(1)} + A_2\mathbf{e}^{(2)} + A_3\mathbf{e}^{(3)}$, где $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$ и $\mathbf{e}^{(3)}$ — взаимные базисные векторы, т. е. векторы, удовлетворяющие соотношениям $\mathbf{e}^{(i)}\mathbf{e}_k = \delta_{ik}$, где δ_{ik} — символ Кронекера. В специальном случае, когда $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ ортогональны, $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$ и $\mathbf{e}^{(3)}$ также ортогональны и соответствующие векторы обеих групп параллельны. В этом случае можно ввести одну группу *естественных* компонент $\overline{A}_i = \sqrt{A_i A^{(i)}}$, которые допускают простую геометрическую интерпретацию: они являются ортогональными проекциями вектора $\mathbf{A}$ на три направления. В случае сферических координат естественные компоненты определяются выражениями (8). Компоненты тензора можно рассматривать аналогичным образом.

Применяя формулы (8) к вектору $\operatorname{rot}\mathbf{A}$, получим

$$\left.\begin{aligned} (\operatorname{rot}\mathbf{A})_r &= \frac{1}{r^2\sin\theta}\left\{\frac{\partial(rA_\varphi\sin\theta)}{\partial\theta} - \frac{\partial(rA_\theta)}{\partial\varphi}\right\}, \\ (\operatorname{rot}\mathbf{A})_\theta &= \frac{1}{r\sin\theta}\left\{\frac{\partial A_r}{\partial\varphi} - \frac{\partial(rA_\varphi\sin\theta)}{\partial r}\right\}, \\ (\operatorname{rot}\mathbf{A})_\varphi &= \frac{1}{r}\left\{\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial\theta}\right\}. \end{aligned}\right\} \tag{9}$$

В сферических координатах уравнения поля (1) примут вид

$$\left.\begin{array}{l}\left.\begin{array}{ll}-k_1E_r=\dfrac{1}{r^2\sin\theta}\left\{\dfrac{\partial(rH_\varphi\sin\theta)}{\partial\theta}-\dfrac{\partial(rH_\theta)}{\partial\varphi}\right\}, & (\alpha)\\ -k_1E_\theta=\dfrac{1}{r\sin\theta}\left\{\dfrac{\partial H_r}{\partial\varphi}-\dfrac{\partial(rH_\varphi\sin\theta)}{\partial r}\right\}, & (\beta)\\ -k_1E_\varphi=\dfrac{1}{r}\left\{\dfrac{\partial(rH_\theta)}{\partial r}-\dfrac{\partial H_r}{\partial\theta}\right\}. & (\gamma)\end{array}\right\}(а)\\ \left.\begin{array}{ll}k_2H_r=\dfrac{1}{r^2\sin\theta}\left\{\dfrac{\partial(rE_\varphi\sin\theta)}{\partial\theta}-\dfrac{\partial(rE_\theta)}{\partial\varphi}\right\}, & (\alpha)\\ k_2H_\theta=\dfrac{1}{r\sin\theta}\left\{\dfrac{\partial E_r}{\partial\varphi}-\dfrac{\partial(rE_\varphi\sin\theta)}{\partial r}\right\}, & (\beta)\\ k_2H_\varphi=\dfrac{1}{r}\left\{\dfrac{\partial(rE_\theta)}{\partial r}-\dfrac{\partial E_r}{\partial\theta}\right\}. & (\gamma)\end{array}\right\}(б)\end{array}\right\}\qquad(10)$$

Граничные условия (5) теперь запишутся следующим образом:

$$\left.\begin{array}{ll}E_\theta^{(I)}=E_\theta^{(II)}, & E_\varphi^{(I)}=E_\varphi^{(II)},\\ H_\theta^{(I)}=H_\theta^{(II)}, & H_\varphi^{(I)}=H_\varphi^{(II)}\end{array}\right\}\text{ для } r=a. \qquad (11)$$

Уравнения (10) вместе с граничными условиями (11) служат основными уравнениями нашей задачи.

Представим решение этих уравнений как суперпозицию двух линейно независимых полей (${}^e\mathbf{E}$, ${}^e\mathbf{H}$) и (${}^m\mathbf{E}$, ${}^m\mathbf{H}$), каждое из которых удовлетворяет уравнениям (10), причем

$${}^eE_r=E_r,\qquad {}^eH_r=0 \qquad (12а)$$

и

$${}^mE_r=0,\qquad {}^mH_r=H_r. \qquad (12б)$$

Нетрудно показать, что такое представление согласуется с нашими уравнениями. Для $H_r={}^eH_r=0$ уравнения (10а, β) и (10а, γ) примут вид

$$\left.\begin{array}{l}k_1{}^eE_\theta=\dfrac{1}{r}\dfrac{\partial}{\partial r}(r{}^eH_\varphi),\\ k_1{}^eE_\varphi=-\dfrac{1}{r}\dfrac{\partial}{\partial r}(r{}^eH_\theta).\end{array}\right\}\qquad(13)$$

Подставляя E_θ и E_φ из этих соотношений в (10б, β) и (10б, γ), получим

$$\left.\begin{array}{ll}\left(\dfrac{\partial^2}{\partial r^2}+k^2\right)(r{}^eH_\theta)=-\dfrac{k_1}{\sin\theta}\dfrac{\partial{}^eE_r}{\partial\varphi}, & (б,\ \beta)\\ \left(\dfrac{\partial^2}{\partial r^2}+k^2\right)(r{}^eH_\varphi)=+k_1\dfrac{\partial{}^eE_r}{\partial\theta}. & (б,\ \gamma)\end{array}\right\}\qquad(14)$$

Уравнения (14) вместе с (10а, α) образуют систему уравнений для ${}^eE_r$, ${}^eH_\theta$ и ${}^eH_\varphi$. Однако физические поля представляют не все решения этой системы, а лишь те, которые удовлетворяют дополнительному условию $\operatorname{div}{}^e\mathbf{H}=0$. Мы ограничимся именно такими решениями. В сферических координатах при нашем предположении, что ${}^eH_r=0$, дополнительное условие запишется следующим образом:

$$\frac{\partial}{\partial\theta}(\sin\theta\,{}^eH_\theta)+\frac{\partial}{\partial\varphi}({}^eH_\varphi)=0. \qquad (15)$$

Оно обеспечивает выполнение оставшегося уравнения (10б, α). Действительно, подставляя (13) в (10б, α), получим

$$0=\frac{1}{k_1^2r^2\sin\theta}\frac{\partial}{\partial r}\left[\frac{\partial}{\partial\theta}(r\sin\theta\,{}^eH_\theta)+\frac{\partial}{\partial\varphi}(r{}^eH_\varphi)\right],$$

а это соотношение выполняется тождественно на основании (15). Совершенно аналогичные рассуждения применимы и ко второму случаю, когда ${}^mE_r=0$.

Решение с нулевым радиальным магнитным полем называется *электрической волной* (или поперечной магнитной волной), а решение с нулевым радиальным электрическим полем — *магнитной волной* (или поперечной электрической волной). Ниже мы покажем, что каждую из волн можно получить из соответствующего скалярного потенциала ${}^e\Pi$ или ${}^m\Pi$, которые известны как *потенциалы Дебая* *).

При ${}^eH_r=0$ из (10б, α) сразу же следует, что ${}^eE_\varphi$ и ${}^eE_\theta$ можно представить как градиенты скаляра

$$ {}^eE_\varphi=\frac{1}{r\sin\theta}\frac{\partial U}{\partial\varphi},\qquad {}^eE_\theta=\frac{1}{r}\frac{\partial U}{\partial\theta}. \tag{16}$$

Если мы положим

$$U=\frac{\partial(r\,{}^e\Pi)}{\partial r}, \tag{17}$$

то из (16) получим

$$ {}^eE_\theta=\frac{1}{r}\frac{\partial^2(r\,{}^e\Pi)}{\partial r\,\partial\theta},\qquad {}^eE_\varphi=\frac{1}{r\sin\theta}\frac{\partial^2(r\,{}^e\Pi)}{\partial r\,\partial\varphi}. \tag{18}$$

Как мы видим, уравнениям (13) удовлетворяют

$$\left.\begin{aligned} {}^eH_\varphi&=k_1\frac{\partial\,{}^e\Pi}{\partial\theta}=\frac{k_1}{r}\frac{\partial(r\,{}^e\Pi)}{\partial\theta},\\ {}^eH_\theta&=-\frac{k_1}{\sin\theta}\frac{\partial\,{}^e\Pi}{\partial\varphi}=-\frac{k_1}{r\sin\theta}\frac{\partial(r\,{}^e\Pi)}{\partial\varphi}. \end{aligned}\right\} \tag{19}$$

Если подставить (19) в (10а, α), то найдем

$$ {}^eE_r=-\frac{1}{r\sin\theta}\left\{\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\,{}^e\Pi}{\partial\theta}\right)+\frac{1}{\sin\theta}\frac{\partial^2\,{}^e\Pi}{\partial\varphi^2}\right\}. \tag{20}$$

Подстановка (19) и (20) в (14) дает два уравнения, первое из которых выражает равенство нулю производной по φ, а второе — производной по θ одного и того же выражения. Поэтому, приравнивая это выражение нулю, можно удовлетворить обоим уравнениям. Таким образом, мы получим

$$\frac{1}{r}\frac{\partial^2(r\,{}^e\Pi)}{\partial r^2}+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\,{}^e\Pi}{\partial\theta}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\,{}^e\Pi}{\partial\varphi^2}+k^2\,{}^e\Pi=0. \tag{21}$$

С помощью (21) выражение (20) можно переписать в виде

$$ {}^eE_r=\frac{\partial^2(r\,{}^e\Pi)}{\partial r^2}+k^2r\,{}^e\Pi. \tag{22}$$

Подставляя (18), (19), (20), (21) и (22) в уравнения (10), легко убедиться, что мы получили решение нашей системы уравнений.

Подобным же образом можно рассмотреть магнитную волну; тогда мы увидим, что эта волна определяется потенциалом ${}^m\Pi$, удовлетворяющим тому же дифференциальному уравнению (21), что ${}^e\Pi$. Полным решением уравнений поля служит сумма двух полей, т. е.

$$\left.\begin{aligned} E_r&={}^eE_r+{}^mE_r=\frac{\partial^2(r\,{}^e\Pi)}{\partial r^2}+k^2r\,{}^e\Pi, && (\alpha)\\ E_\theta&={}^eE_\theta+{}^mE_\theta=\frac{1}{r}\frac{\partial^2(r\,{}^e\Pi)}{\partial r\,\partial\theta}+k_2\frac{1}{r\sin\theta}\frac{\partial(r\,{}^m\Pi)}{\partial\varphi}, && (\beta)\\ E_\varphi&={}^eE_\varphi+{}^mE_\varphi=\frac{1}{r\sin\theta}\frac{\partial^2(r\,{}^e\Pi)}{\partial r\,\partial\varphi}-k_2\frac{1}{r}\frac{\partial(r\,{}^m\Pi)}{\partial\theta}, && (\gamma) \end{aligned}\right\} \tag{23а}$$

$$\left.\begin{aligned} H_r&={}^mH_r+{}^eH_r=k^2r\,{}^m\Pi+\frac{\partial^2(r\,{}^m\Pi)}{\partial r^2}, && (\alpha)\\ H_\theta&={}^mH_\theta+{}^eH_\theta=-k_1\frac{1}{r\sin\theta}\frac{\partial(r\,{}^e\Pi)}{\partial\varphi}+\frac{1}{r}\frac{\partial(r\,{}^m\Pi)}{\partial r\,\partial\theta}, && (\beta)\\ H_\varphi&={}^mH_\varphi+{}^eH_\varphi=k_1\frac{1}{r}\frac{\partial(r\,{}^e\Pi)}{\partial\theta}+\frac{1}{r\sin\theta}\frac{\partial^2(r\,{}^m\Pi)}{\partial r\,\partial\varphi}. && (\gamma) \end{aligned}\right\} \tag{23б}$$

*) Если $\mathbf{r}$ — радиус-вектор из начала координат, то ${}^e\Pi\mathbf{r}$ и ${}^m\Pi\mathbf{r}$ — *векторы Герца*, всюду имеющие радиальное направление, о которых говорилось в п. 2.2.2 (см. [24], раздел, написанный А. Зоммерфельдом, а также [25]).

Оба потенциала ${}^e\Pi$ и ${}^m\Pi$ служат решениями дифференциального уравнения (21), которое представляет собой не что иное, как волновое уравнение

$$\nabla^2\Pi + k^2\Pi = 0,$$

записанное в сферических координатах. Для того чтобы компоненты E_θ, E_φ, H_θ и H_φ были непрерывны на сферической поверхности $r = a$, очевидно, достаточно, чтобы на этой поверхности были *непрерывны* следующие четыре величины:

$$k_1 r\,{}^e\Pi, \qquad k_2 r\,{}^m\Pi, \qquad \frac{\partial}{\partial r}(r\,{}^e\Pi), \qquad \frac{\partial}{\partial r}(r\,{}^m\Pi). \tag{24}$$

Иными словами наши граничные условия также разделяются на независимые условия для ${}^e\Pi$ и ${}^m\Pi$. Таким образом, рассматриваемая нами дифракционная задача сводится к проблеме получения двух взаимно независимых решений волнового уравнения с заданными граничными условиями.

б. Разложение в ряды компонент полей. Вначале мы представим решение волновых уравнений в виде ряда с неопределенными коэффициентами, причем каждый член этого ряда будет каким-то частным решением уравнения. Затем, используя граничные условия, определим коэффициенты.

Будем искать решение в виде

$$\Pi = R(r)\,\Theta(\theta)\,\Phi(\varphi). \tag{25}$$

Функции R, Θ и Φ, как легко проверить прямой подстановкой в (21), должны удовлетворять обычным дифференциальным уравнениям

$$\frac{d^2(rR)}{dr^2} + \left(k^2 - \frac{\alpha}{r^2}\right) rR = 0, \tag{26а}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \left(\alpha - \frac{\beta}{\sin^2\theta}\right)\Theta = 0, \tag{26б}$$

$$\frac{d^2\Phi}{d\varphi^2} + \beta\Phi = 0, \tag{26в}$$

где α и β — постоянные интегрирования.

Так как поле **E**, **H** является однозначной функцией координат, функция Π также должна быть однозначной. Это требование налагает определенные условия на Θ и Φ.

Для каждого из уравнений (26) можно записать общее решение. Для (26в) оно имеет вид

$$a\cos(\sqrt{\beta}\varphi) + b\sin(\sqrt{\beta}\varphi).$$

Условие однозначности требует, чтобы

$$\beta = m^2 \qquad (m \text{ — целое}). \tag{27}$$

Следовательно, однозначное решение (26в) имеет вид

$$\Phi = a_m\cos(m\varphi) + b_m\sin(m\varphi). \tag{28}$$

Уравнение (26б) — это хорошо известное уравнение сферических гармоник. Необходимым и достаточным условием однозначности его решения является

$$\alpha = l(l+1), \qquad (l > |m|, \text{ целое}). \tag{29}$$

Подставим β из (27) в (26б) и введем новую переменную

$$\xi = \cos\theta. \tag{30}$$

Тогда (26б) преобразуется к виду (см. [26])

$$\frac{d}{d\xi}\left\{(1-\xi^2)\frac{d\Theta}{d\xi}\right\} + \left\{l(l+1) - \frac{m^2}{1-\xi^2}\right\}\Theta = 0. \tag{31}$$

Решениями последнего уравнения служат присоединенные функции Лежандра, т. е.

$$\Theta = P_l^{(m)}(\xi) = P_l^{(m)}(\cos\theta). \tag{32}$$

Эти функции тождественно равны нулю, если $|m|>l$, поэтому для каждого l имеется $2l+1$ таких функций, а именно функции с

$$m=-l,\ -l+1,\ \ldots,\ l-1,\ l.$$

Чтобы проинтегрировать последнее уравнение (26а), положим

$$kr=\rho,\qquad R_l(r)=\frac{1}{\sqrt{\rho}}Z(\rho). \tag{33}$$

Тогда мы получим уравнение Бесселя (см. [26])

$$\frac{d^2Z}{d\rho^2}+\frac{1}{\rho}\frac{dZ}{d\rho}+\left\{1-\frac{\left(l+\frac{1}{2}\right)^2}{\rho^2}\right\}Z=0. \tag{34}$$

Решением этого уравнения является цилиндрическая функция общего вида $Z_l=Z_{l+1/2}(\rho)$ порядка $l+1/2$, так что решение (26а) запишется следующим образом:

$$R=\frac{1}{\sqrt{kr}}Z_{l+1/2}(kr). \tag{35}$$

Каждую цилиндрическую функцию можно выразить в виде линейной комбинации двух цилиндрических функций обычного типа, например функций Бесселя $J_{l+1/2}(\rho)$ и функций Неймана $N_{l+1/2}(\rho)$. Для нашей цели удобно применить функции *)

$$\psi_l(\rho)=\sqrt{\frac{\pi\rho}{2}}J_{l+1/2}(\rho),\qquad \chi_l(\rho)=-\sqrt{\frac{\pi\rho}{2}}N_{l+1/2}(\rho). \tag{36}$$

Функции $\psi_l(\rho)$ регулярны в каждой конечной области плоскости ρ, включая начало координат; функции $\chi_l(\rho)$ имеют особенности в начале координат $\rho=0$, где они обращаются в бесконечность. Поэтому для представления волны внутри сферы нужно использовать функции $\psi_l(\rho)$, а не $\chi_l(\rho)$.

Общее решение уравнения (26а) можно записать в виде

$$rR=c_l\psi_l(kr)+d_l\chi_l(kr). \tag{37}$$

В частности, для $c_l=1$, $d_l=-i$ мы получим

$$rR=\zeta_l^{(1)}(kr), \tag{38}$$

где

$$\zeta_l^{(1)}(\rho)=\psi_l(\rho)-i\chi_l(\rho)=\sqrt{\frac{\pi\rho}{2}}H^{(1)}_{l+1/2}(\rho). \tag{39}$$

Здесь $H^{(1)}$ — одна из функций Ханкеля **). Функции Ханкеля отличаются от других цилиндрических функций тем, что в комплексной плоскости они равны нулю на бесконечности. Для функции с индексом 1 это справедливо в полуплоскости с положительным значением мнимой части ρ. Таким образом, именно такая функция оказывается подходящей для представления рассеянной волны.

Согласно (25) частное решение $\Pi_l^{(m)}$ получается при перемножении функций, определяемых выражениями (28), (32) и (37). Общее решение нашего

*) В литературе существует несколько немного различающихся определений функций ψ и χ (см. [27] или [26]). Ниже мы приведем некоторые формулы, относящиеся к функциям ψ_l.

**) Эта формула непосредственно следует из хорошо известного соотношения между функциями Бесселя, Неймана и Ханкеля, а именно $J_p+iN_p=H_p^{(1)}$. Аналогичное соотношение имеется и для второй функции Ханкеля: $J_p-iN_p=H_p^{(2)}$. Приведенные формулы аналогичны выражениям, связывающим экспоненциальные функции $\exp(i\rho)$ и $\exp(-i\rho)$ с $\cos\rho$ и $\sin\rho$.

волнового уравнения равно

$$r\Pi = r\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\Pi_l^{(m)} =$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\{c_l\psi_l(kr)+d_l\chi_l(kr)\}\{P_l^{(m)}(\cos\theta)\}\{a_m\cos(m\varphi)+b_m\sin(m\varphi)\}, \tag{40}$$

где a_m, b_m, c_l и d_l — произвольные постоянные.

Эти постоянные необходимо выбрать так, чтобы удовлетворить граничным условиям. Для этого нужно выразить потенциалы ${}^e\Pi^{(i)}$ и ${}^m\Pi^{(i)}$ падающей волны в виде рядов, аналогичных ряду (40). Запишем вначале выражение (4) для падающей волны в сферических координатах в соответствии с (8):

$$\left.\begin{aligned} E_r^{(i)} &= \exp(ik^{(\mathrm{I})}r\cos\theta)\sin\theta\cos\varphi, & H_r^{(i)} &= \frac{ik^{(\mathrm{I})}}{k_2^{(\mathrm{I})}}\exp(ik^{(\mathrm{I})}r\cos\theta)\sin\theta\sin\varphi,\\ E_\theta^{(i)} &= \exp(ik^{(\mathrm{I})}r\cos\theta)\cos\theta\cos\varphi, & H_\theta^{(i)} &= \frac{ik^{(\mathrm{I})}}{k_2^{(\mathrm{I})}}\exp(ik^{(\mathrm{I})}r\cos\theta)\cos\theta\sin\varphi,\\ E_\varphi^{(i)} &= -\exp(ik^{(\mathrm{I})}r\cos\theta)\sin\varphi, & H_\varphi^{(i)} &= \frac{ik^{(\mathrm{I})}}{k_2^{(\mathrm{I})}}\exp(ik^{(\mathrm{I})}r\cos\theta)\cos\varphi. \end{aligned}\right\} \tag{41}$$

Для определения потенциалов ${}^e\Pi^{(i)}$ или ${}^m\Pi^{(i)}$ необходимо лишь воспользоваться одним из уравнений (23). Первое из них дает

$$\exp(ik^{(\mathrm{I})}r\cos\theta)\sin\theta\cos\varphi = \frac{\partial^2(r^e\Pi)}{\partial r^2} + (k^{(\mathrm{I})})^2 r^e\Pi^{(i)}. \tag{42}$$

Первый множитель в левой части этого уравнения можно выразить в виде следующего ряда полиномов Лежандра (формула Бауэра, см. (9.4.9)):

$$\exp(ik^{(\mathrm{I})}r\cos\theta) = \sum_{l=0}^{\infty} i^l(2l+1)\frac{\psi_l(k^{(\mathrm{I})}r)}{k^{(\mathrm{I})}r}P_l(\cos\theta). \tag{43}$$

Выполняются также тождества

$$\exp(ik^{(\mathrm{I})}r\cos\theta)\sin\theta \equiv -\frac{1}{ik^{(\mathrm{I})}r}\frac{\partial}{\partial\theta}(\exp[ik^{(\mathrm{I})}r\cos\theta]), \tag{44}$$

$$\frac{\partial}{\partial\theta}P_l(\cos\theta) \equiv -P_l^{(1)}(\cos\theta);\quad P_0^{(1)}(\cos\theta)\equiv 0. \tag{45}$$

Используя эти соотношения, можно переписать левую часть (42) в виде

$$\exp(ik^{(\mathrm{I})}r\cos\theta)\sin\theta\cos\varphi = \frac{1}{(k^{(\mathrm{I})}r)^2}\sum_{l=1}^{\infty} i^{l-1}(2l+1)\psi_l(k^{(\mathrm{I})}r)P_l^{(1)}(\cos\theta)\cos\varphi. \tag{46}$$

Выберем в качестве пробного решения уравнения (42) ряд такого же вида

$$r^e\Pi^{(i)} = \frac{1}{(k^{(\mathrm{I})})^2}\sum_{l=1}^{\infty}\alpha_l\psi_l(k^{(\mathrm{I})}r)P_l^{(1)}(\cos\theta)\cos\varphi. \tag{47}$$

Подставляя (46) и (47) в уравнение (42) и сравнивая коэффициенты, получим соотношение

$$\alpha_l\left\{(k^{(\mathrm{I})})^2\psi_l(k^{(\mathrm{I})}r) + \frac{\partial^2\psi_l(k^{(\mathrm{I})}r)}{\partial r^2}\right\} = i^{l-1}(2l+1)\frac{\psi_l(k^{(\mathrm{I})}r)}{r^2}. \tag{48}$$

Из (37) (при $c_l = 1$ и $d_l = 0$) следует, что

$$\psi_l(k^{(\mathrm{I})}r) = rR \tag{49}$$

является решением уравнения (26а):

$$\frac{d^2\psi_l}{dr^2} + \left((k^{(\mathrm{I})})^2 - \frac{\alpha}{r^2}\right)\psi_l = 0 \tag{50}$$

при условии, что $\alpha = l(l+1)$ (см. (29)). Сравнивая (50) с (48), мы видим, что

$$\alpha_l = i^{l-1}\frac{2l+1}{l(l+1)}. \qquad (51)$$

Вычисления, связанные с магнитным потенциалом ${}^m\Pi^{(i)}$, аналогичны Таким образом, мы получим следующие выражения для потенциалов падающей волны:

$$\left.\begin{aligned} r\,{}^e\Pi^{(i)} &= \frac{1}{(k^{(\mathrm{I})})^2}\sum_{l=1}^{\infty} i^{l-1}\frac{2l+1}{l(l+1)}\psi_l(k^{(\mathrm{I})}r)\,P_l^{(1)}(\cos\theta)\cos\varphi, \quad (а)\\ r\,{}^m\Pi^{(i)} &= \frac{1}{(k^{(\mathrm{I})})^2}\sum_{l=1}^{\infty} i^{l}\frac{k^{(\mathrm{I})}}{k_2^{(\mathrm{I})}}\frac{2l+1}{l(l+1)}\psi_l(k^{(\mathrm{I})}r)\,P_l^{(1)}(\cos\theta)\sin\varphi. \quad (б)\end{aligned}\right\} \qquad (52)$$

Мы выразили оба потенциала в виде рядов, аналогичных ряду (40), и теперь легко определить неизвестные постоянные.

Запишем граничные условия (24) полностью

$$\left.\begin{aligned} \frac{\partial}{\partial r}\{r({}^e\Pi^{(i)}+{}^e\Pi^{(s)})\}_{r=a} &= \frac{\partial}{\partial r}\{r\,{}^e\Pi^{(w)}\}_{r=a}, \quad (а)\\ \frac{\partial}{\partial r}\{r({}^m\Pi^{(i)}+{}^m\Pi^{(s)})\}_{r=a} &= \frac{\partial}{\partial r}\{r\,{}^m\Pi^{(w)}\}_{r=a}, \quad (б)\\ k_1^{(\mathrm{I})}\{r({}^e\Pi^{(i)}+{}^e\Pi^{(s)})\}_{r=a} &= k_1^{(\mathrm{II})}\{r\,{}^e\Pi^{(w)}\}_{r=a}, \quad (в)\\ k_2^{(\mathrm{I})}\{r({}^m\Pi^{(i)}+{}^m\Pi^{(s)})\}_{r=a} &= k_2^{(\mathrm{II})}\{r\,{}^m\Pi^{(w)}\}_{r=a}. \quad (г)\end{aligned}\right\} \qquad (53)$$

Из соотношений (52) следует, что эти уравнения могут удовлетворяться лишь в том случае, если ряды, аналогичные ряду (40), для неизвестных потенциалов $\Pi^{(s)}$ и $\Pi^{(w)}$ содержат только члены с $m = 1$ и если, кроме того, для магнитного потенциала

$$a_1 = 0,$$

а для электрического —

$$b_1 = 0.$$

Мы уже отмечали, что для представления $\Pi^{(w)}$ пригодны лишь функции ψ_l, поскольку они остаются конечными в начале координат, тогда как функции χ_l обращаются там в бесконечность. Следовательно, можно написать

$$\left.\begin{aligned} r\,{}^e\Pi^{(w)} &= \frac{1}{(k^{(\mathrm{II})})^2}\sum_{l=1}^{\infty} {}^eA_l\psi_l(k^{(\mathrm{II})}r)\,P_l^{(1)}(\cos\theta)\cos\varphi, \quad (а)\\ r\,{}^m\Pi^{(w)} &= \frac{i}{k^{(\mathrm{II})}k_2^{(\mathrm{II})}}\sum_{l=1}^{\infty} {}^mA_l\psi_l(k^{(\mathrm{II})}r)\,P_l^{(1)}(\cos\theta)\sin\varphi. \quad (б)\end{aligned}\right\} \qquad (54)$$

Мы видели также, что рассеянную волну можно выразить с помощью функций $\zeta_l^{(1)} = \psi_l - i\chi_l$, которые получаются при умножении функции Ханкеля $H^{(1)}$ на $\sqrt{\pi\rho/2}$ (см. (39)). При больших значениях ρ функция $H^{(1)}$ ведет себя как $e^{i\rho}/\sqrt{\rho}$, т. е. $\zeta_l^{(1)}$ — как $e^{i\rho}$ и $R = \zeta_l^{(1)}(k^{(\mathrm{I})}r)/r$ — как $\exp(ik^{(\mathrm{I})}r)/r$. Таким образом, на больших расстояниях от сферы рассеянная волна сферична с центром в начале координат $r = 0$. Поэтому положим

$$\left.\begin{aligned} r\,{}^e\Pi^{(s)} &= \frac{1}{(k^{(\mathrm{I})})^2}\sum_{l=1}^{\infty} {}^eB_l\zeta_l^{(1)}(k^{(\mathrm{I})}r)\,P_l^{(1)}(\cos\theta)\cos\varphi, \quad (в)\\ r\,{}^m\Pi^{(s)} &= \frac{i}{k^{(\mathrm{I})}k_2^{(\mathrm{I})}}\sum_{l=1}^{\infty} {}^mB_l\zeta_l^{(1)}(k^{(\mathrm{I})}r)\,P_l^{(1)}(\cos\theta)\sin\varphi. \quad (г)\end{aligned}\right\} \qquad (55)$$

Если подставить выражения (52), (54) и (55) в граничные условия (53), мы получим следующие линейные соотношения между коэффициентами *) ${}^eA_l$, ${}^mA_l$, ${}^eB_l$ и ${}^mB_l$:

$$\left.\begin{aligned}
&{}^eB_l \frac{1}{k^{(I)}} \zeta_l^{(1)\prime}(k^{(I)}a) + \frac{1}{k^{(I)}} i^{l-1} \frac{2l+1}{l(l+1)} \psi_l'(k^{(I)}a) = \frac{1}{k^{(II)}}\, {}^eA_l \psi_l'(k^{(II)}a),\\
&{}^mB_l \frac{1}{k_2^{(I)}} \zeta^{(1)\prime}(k^{(I)}a) + \frac{1}{k_2^{(I)}} i^{l-1} \frac{2l+1}{l(l+1)} \psi_l'(k^{(I)}a) = \frac{1}{k_2^{(II)}}\, {}^mA_l \psi_l'(k^{(II)}a),\\
&{}^eB_l \frac{1}{k_2^{(I)}} \zeta_l^{(1)}(k^{(I)}a) + \frac{1}{k_2^{(I)}} i^{l-1} \frac{2l+1}{l(l+1)} \psi_l(k^{(I)}a) = \frac{1}{k_2^{(II)}}\, {}^eA_l \psi_l(k^{(II)}a),\\
&{}^mB_l \frac{1}{k^{(I)}} \zeta_l^{(1)}(k^{(I)}a) + \frac{1}{k^{(I)}} i^{l-1} \frac{2l+1}{l(l+1)} \psi_l(k^{(I)}a) = \frac{1}{k^{(II)}}\, {}^mA_l \psi_l(k^{(II)}a).
\end{aligned}\right\} \qquad (56)$$

Нас интересуют только коэффициенты ${}^eB_l$ и ${}^mB_l$, которые характеризуют рассеянную волну. Исключая ${}^eA_l$ и ${}^mA_l$, получим

$$\left.\begin{aligned}
{}^eB_l &= i^{l+1} \frac{2l+1}{l(l+1)} \frac{k_2^{(I)} k^{(II)} \psi_l'(k^{(I)}a)\, \psi_l(k^{(II)}a) - k_2^{(II)} k^{(I)} \psi_l'(k^{(II)}a)\, \psi_l(k^{(I)}a)}{k_2^{(I)} k^{(II)} \zeta_l^{(1)\prime}(k^{(I)}a)\, \psi_l(k^{(II)}a) - k_2^{(II)} k^{(I)} \psi_l'(k^{(II)}a)\, \zeta^{(1)}(k^{(I)}a)}, \quad (\text{а})\\
{}^mB_l &= i^{l+1} \frac{2l+1}{l(l+1)} \frac{k_2^{(I)} k^{(II)} \psi_l(k^{(I)}a)\, \psi_l'(k^{(II)}a) - k_2^{(II)} k^{(I)} \psi_l'(k^{(I)}a)\, \psi_l(k^{(II)}a)}{k_2^{(I)} k^{(II)} \zeta_l^{(1)}(k^{(I)}a)\, \psi_l'(k^{(II)}a) - k_2^{(II)} k^{(I)} \zeta_l^{(1)\prime}(k^{(I)}a)\, \psi_l(k^{(II)}a)}. \quad (\text{б})
\end{aligned}\right\} \qquad (57)$$

Наконец, при подстановке (55) в (23) найдем для компонент векторов поля рассеянной волны

$$\left.\begin{aligned}
E_r^{(s)} &= \frac{1}{(k^{(I)})^2} \frac{\cos\varphi}{r^2} \sum_{l=1}^{\infty} l(l+1)\, {}^eB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)}(\cos\theta),\\
E_\theta^{(s)} &= -\frac{1}{k^{(I)}} \frac{\cos\varphi}{r} \sum_{l=1}^{\infty} \Big\{ {}^eB_l \zeta_l^{(1)\prime}(k^{(I)}r)\, P_l^{(1)\prime}(\cos\theta) \sin\theta -\\
&\qquad - i\, {}^mB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)}(\cos\theta) \frac{1}{\sin\theta} \Big\},\\
E_\varphi^{(s)} &= -\frac{1}{k^{(I)}} \frac{\sin\varphi}{r} \sum_{l=1}^{\infty} \Big\{ {}^eB_l \zeta_l^{(1)\prime}(k^{(I)}r)\, P_l^{(1)}(\cos\theta) \frac{1}{\sin\theta} -\\
&\qquad - i\, {}^mB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)\prime}(\cos\theta) \sin\theta \Big\},\\
H_r^{(s)} &= \frac{i}{k^{(I)} k_2^{(I)}} \frac{\sin\varphi}{r^2} \sum_{l=1}^{\infty} l(l+1)\, {}^mB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)}(\cos\theta),\\
H_\theta^{(s)} &= -\frac{1}{k_2^{(I)}} \frac{\sin\varphi}{r} \sum_{l=1}^{\infty} \Big\{ {}^eB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)}(\cos\theta) \frac{1}{\sin\theta} +\\
&\qquad + i\, {}^mB_l \zeta_l^{(1)\prime}(k^{(I)}r)\, P_l^{(1)\prime}(\cos\theta) \sin\theta \Big\},\\
H_\varphi^{(s)} &= \frac{1}{k_2^{(I)}} \frac{\cos\varphi}{r} \sum_{l=1}^{\infty} \Big\{ {}^eB_l \zeta_l^{(1)}(k^{(I)}r)\, P_l^{(1)\prime}(\cos\theta) \sin\theta +\\
&\qquad + i\, {}^mB_l \zeta_l^{(1)\prime}(k^{(I)}r)\, P_l^{(1)}(\cos\theta) \frac{1}{\sin\theta} \Big\}.
\end{aligned}\right\} \qquad (58)$$

*) Штрих у функций ψ_l, ζ_l и $P_l^{(1)}$ означает дифференцирование по их аргументу.

Этим завершается формальное решение нашей граничной задачи. Мы не будем здесь заниматься вопросами существования и сходимости полученного решения.

Полезно напомнить значение различных постоянных. Так как, по предположению, сферу окружает непроводящая среда, то $\sigma^{(I)}=0$. Из соотношений (2) следует

$$\left.\begin{aligned}
&k_1^{(I)}=\frac{i\omega}{c}\varepsilon^{(I)}=i\frac{2\pi}{\lambda_0}\varepsilon^{(I)},\quad k_2^{(I)}=\frac{i\omega}{c}=i\frac{2\pi}{\lambda_0}, && \text{(а)}\\
&k^{(I)}=\sqrt{-k_1^{(I)}k_2^{(I)}}=\frac{2\pi}{\lambda_0}\sqrt{\varepsilon^{(I)}}=\frac{2\pi}{\lambda^{(I)}}, && \text{(б)}\\
&k_1^{(II)}=\frac{i\omega}{c}\left(\varepsilon^{(II)}+i\frac{4\pi\sigma}{\omega}\right)=i\frac{2\pi}{\lambda_0}\left(\varepsilon^{(II)}+i\frac{4\pi\sigma}{\omega}\right),\quad k_2^{(II)}=\frac{i\omega}{c}=i\frac{2\pi}{\lambda_0}, && \text{(в)}\\
&k^{(II)}=\sqrt{-k_1^{(II)}k_2^{(II)}}=\frac{2\pi}{\lambda_0}\sqrt{\varepsilon^{(II)}+i\frac{4\pi\sigma}{\omega}}, && \text{(г)}
\end{aligned}\right\}\quad(59)$$

здесь λ_0 — длина волны света в вакууме, $\lambda^{(I)}$ — длина волны в среде, окружающей сферу, и проводимость сферы $\sigma^{(II)}$ обозначена через σ.

Удобно также ввести комплексный показатель преломления сферы относительно окружающей среды, который понадобится нам в дальнейшем. Обозначая этот показатель через $\hat{n}$, получим

$$\hat{n}^2=\frac{(\hat{n}^{(II)})^2}{(n^{(I)})^2}=\frac{(k^{(II)})^2}{(k^{(I)})^2}=\frac{\varepsilon^{(II)}}{\varepsilon^{(I)}}+i\frac{4\pi\sigma}{\omega\varepsilon^{(I)}}=\frac{k_1^{(II)}}{k_1^{(I)}}. \qquad (60)$$

Введем, кроме того, безразмерный параметр q, определяемый выражением

$$q=\frac{2\pi}{\lambda^{(I)}}a, \qquad (61)$$

т. е. величина q равна произведению 2π на отношение радиуса сферы к длине волны света во внешней среде. Тогда, используя соотношение

$$\frac{k^{(II)}k_2^{(I)}}{k^{(I)}k_2^{(II)}}=\hat{n},$$

можно выразить коэффициенты (57) в виде

$$\left.\begin{aligned}
&{}^eB_l=i^{l+1}\frac{2l+1}{l(l+1)}\,\frac{\hat{n}\psi_l'(q)\psi_l(\hat{n}q)-\psi_l(q)\psi_l'(\hat{n}q)}{\hat{n}\zeta_l^{(1)\prime}(q)\psi_l(\hat{n}q)-\zeta_l^{(1)}(q)\psi_l'(\hat{n}q)}, && \text{(а)}\\
&{}^mB_l=i^{l+1}\frac{2l+1}{l(l+1)}\,\frac{\hat{n}\psi_l(q)\psi_l'(\hat{n}q)-\psi_l'(q)\psi_l(\hat{n}q)}{\hat{n}\zeta_l^{(1)}(q)\psi_l'(\hat{n}q)-\zeta_l^{(1)\prime}(q)\psi_l(\hat{n}q)}. && \text{(б)}
\end{aligned}\right\}\quad(62)$$

Эти формулы принимают особенно простую форму, когда либо диэлектрическая проницаемость, либо проводимость сферы высоки и вместе с тем радиус сферы не слишком мал. В этом случае $|\hat{n}|\gg 1$, $|\hat{n}q|\gg 1$, и выражения (62) сводятся к

$$\left.\begin{aligned}
&{}^eB_l=i^{l+1}\frac{2l+1}{l(l+1)}\,\frac{\psi_l'(q)}{\zeta_l^{(1)\prime}(q)}, && \text{(а)}\\
&{}^mB_l=i^{l+1}\frac{2l+1}{l(l+1)}\,\frac{\psi_l(q)}{\zeta_l^{(1)}(q)}. && \text{(б)}
\end{aligned}\right\}\quad(63)$$

Хотя это приближение не представляет интереса в оптике, его важность для радиодиапазона несомненна. Оно имеет и исторический интерес, поскольку первые теории относились к такому предельному случаю [28].

в. Сводка формул, относящихся к присоединенным функциям Лежандра и цилиндрическим функциям. Прежде чем переходить к дальнейшему обсуждению, удобно привести здесь некоторые формулы, относящиеся к сферическим гармоникам и цилиндрическим функциям.

Присоединенные функции Лежандра

Полиномы Лежандра (в которых аргументом служит $\cos\theta$) имеют вид

$$P_l(\cos\theta)=\sum_{m=0}^{[l/2]}(-1)^m\frac{(2l-2m)!}{2^l m!(l-m)!(l-2m)!}(\cos\theta)^{l-2m}. \tag{64}$$

Присоединенные функции Лежандра первого рода определяются выражением *)

$$P_l^{(m)}(\cos\theta)=(\sin\theta)^m\frac{d^mP_l(\cos\theta)}{d(\cos\theta)^m}. \tag{65}$$

Нам понадобятся также соотношения

$$\left.\begin{aligned}P_l^{(1)}(\cos\theta)&=\frac{1}{\sin\theta}\{P_{l-1}(\cos\theta)-\cos\theta\,P_l(\cos\theta)\},\\ P_l^{(1)\prime}(\cos\theta)&=\frac{\cos\theta}{\sin^2\theta}\left\{P_l^{(1)}(\cos\theta)-l(l+1)\frac{P_l(\cos\theta)}{\sin\theta}\right\}.\end{aligned}\right\} \tag{66}$$

Для больших значений l справедливо асимптотическое приближение

$$P_l(\cos\theta)\approx\sqrt{\frac{2}{l\pi\sin\theta}}\sin\left[\left(l+\frac{1}{2}\right)\theta+\frac{\pi}{4}\right]. \tag{67}$$

Цилиндрические функции

I. *При малых значениях аргумента* x функцию $\psi_l(x)$ можно разложить в ряд

$$\psi_l(x)=\frac{x^{l+1}}{1\cdot3\cdot\ldots\cdot(2l+1)}f_l(x), \tag{68}$$

где

$$f_l(x)=1-\frac{2}{2l+3}\left(\frac{x}{2}\right)^2+\ldots \tag{69}$$

Для $\zeta_l^{(1)}(x)$ справедливо разложение

$$\zeta_l^{(1)}(x)=-i\,\frac{1\cdot3\ldots(2l-1)}{x^l}e^{ix}\{h_l(x)-ixg_l(x)\}, \tag{70}$$

где $h_l(x)$ и $g_l(x)$ — степенные ряды, в которых первый член равен единице, а второй квадратичен относительно x. Аналогично функции $\psi_l'(x)$ и $\zeta_l^{(1)\prime}(x)$ можно выразить в виде

$$\psi_l'(x)=\frac{(l+1)x^l}{1\cdot3\ldots(2l+1)}f_l^+(x), \tag{71}$$

$$\zeta_l^{(1)\prime}(x)=il\,\frac{1\cdot3\ldots(2l-1)}{x^{l+1}}e^{ix}\{h_l^+(x)-ig_l^+(x)\}, \tag{72}$$

где $f_l^+(x)$, $h_l^+(x)$ и $g_l^+(x)$ — степенные ряды того же типа, что и прежде.

II. *При больших значениях аргумента* x и при условии, что l мало по сравнению с $|x|$, можно использовать следующие асимптотические формулы:

$$\psi_l(x)\approx\frac{1}{2}\{i^{l+1}\exp(-ix)+(-i)^{l+1}\exp(ix)\}; \tag{73}$$

$$\zeta_l^{(1)}(x)\approx(-i)^{l+1}\exp(ix), \tag{74}$$

и

$$\psi_l'(x)\approx\frac{1}{2}\{i^l\exp(-ix)+(-i)^l\exp(ix)\}, \tag{75}$$

$$\zeta_l^{(1)\prime}(x)\approx(-i)^l\exp(ix). \tag{76}$$

---

*) Иногда используется определение, отличающееся от настоящего множителем $(-1)^m$.

Для вещественных значений x функции $\psi_l(x)$ и $\psi_l'(x)$ тоже вещественны, т. е.

$$\psi_l(x) \approx \sin\left(x - \frac{l\pi}{2}\right), \tag{77}$$

$$\psi_l'(x) \approx \cos\left(x - \frac{l\pi}{2}\right). \tag{78}$$

13.5.2. Некоторые следствия из формул Ми. *а. Парциальные волны.* Из выражений (58) видно, что амплитуды радиальных компонент $E_r^{(s)}$ и $H_r^{(s)}$ рассеянной волны уменьшаются обратно пропорционально квадрату расстояния от рассеивающего центра, тогда как амплитуды остальных компонент уменьшаются более медленно, обратно пропорционально первой степени этого расстояния. На достаточно больших расстояниях ($r \gg \lambda$) в *радиационной*, или *волновой, зоне* радиальными компонентами можно пренебречь по сравнению с тангенциальными, т. е. в этой области волны *поперечны*.

Полученные формулы показывают, что рассеянная волна состоит из сферических гармоник разных порядков. Эти гармоники называются *парциальными волнами*, и их амплитуда определяется абсолютными значениями комплексных коэффициентов ${}^eB_l$ и ${}^mB_l$, которые зависят от природы обеих сред и от отношения радиуса рассеивающей сферы к длине волны падающего света.

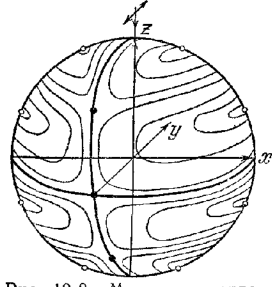


Рис. 13.8 Магнитные силовые линии четвертой электрической парциальной волны.

Каждая парциальная волна состоит из электрической части с амплитудой ${}^eB_l$ и магнитной части с амплитудой ${}^mB_l$. Магнитные силовые линии электрической парциальной волны и электрические силовые линии магнитной парциальной волны целиком лежат на концентрических сферических поверхностях, так как для первой волны $H_r^{(s)} = 0$, для второй — $E_r^{(s)} = 0$.

Рассмотрим произвольную парциальную волну, например l-ю электрическую волну. Мы видим, что соответствующие компоненты $E_\theta^{(s)}$ и $H_\varphi^{(s)}$ равны нулю в точках, где обращаются в нуль либо $\cos\varphi$, либо $P_l^{(1)\prime}(\cos\theta)\sin\theta$. Аналогично $E_\varphi^{(s)}$ и $H_\theta^{(s)}$ исчезают там, где либо $\sin\varphi$, либо $P_l^{(1)}(\cos\theta)/\sin\theta$ равны нулю. Внутри интервала $0 \leqslant \theta \leqslant \pi$ функция $P_l^{(1)\prime}(\cos\theta)$ обращается в нуль l раз, а функция $P_l^{(1)}(\cos\theta)/\sin\theta$ обращается в нуль $l-1$ раз и отличается от нуля при $\theta = 0$ или π. При $\varphi = \pm\pi/2$ все компоненты поля обращаются в нуль $2l$ раз, т. е. всего $(4l-2)$ раза. Поскольку магнитные силовые линии должны быть замкнутыми кривыми и, как мы видели, они целиком расположены на концентрических сферических поверхностях с центром в начале координат, то ясно, что каждая из $2l$ нулевых точек на окружности с $\varphi = 0$ или π оказывается центром семейства таких линий, а каждая из $2(l-1)$ нулевых точек на окружности с $\varphi = \pm\pi/2$ будет нейтральной точкой. Силовые линии обходят эти нейтральные точки, подобно тому, как два семейства равнобочных гипербол с общими асимптотами обходят их общий центр.

На рис. 13.8 показаны магнитные силовые линии четвертой электрической парциальной волны. Отчетливо видны две группы точек; точки первой группы лежат в плоскости xz (плоскость рисунка), точки второй — в плоскости yz.

На рис. 13.9, *а* для первых четырех электрических парциальных волн приведены проекции на плоскость yz их магнитных силовых линий, которые расположены на одной из полусфер, находящихся с любой стороны плоскости yz. В волновой зоне электрические силовые линии ортогональны к магнитным линиям, так как, согласно (58), для каждой парциальной волны (электрической

или магнитной) справедливо соотношение

$$E_\theta^{(s)}H_\theta^{(s)}+E_\varphi^{(s)}H_\varphi^{(s)}=0. \qquad (79)$$

Проекции электрических силовых линий на плоскость xz показаны на рис. 13 9, *б*.

Аналогичные результаты получаются и для магнитных парциальных волн, за исключением того, что $\cos\varphi$ и $\sin\varphi$ меняются местами. Соответствующие проекции электрических силовых линий, расположенных на сфере, для магнитных парциальных волн легко найти, поворачивая рисунки на угол 90° вокруг оси z.

б. Предельные случаи Ниже мы исследуем относительную роль различных парциальных волн Общий случай ($\hat{n}$ и q произвольны) не поддается простому аналитическому рассмотрению, и поэтому мы подробно обсудим лишь два предельных случая, а именно, когда радиус сферы велик по сравнению с длиной волны ($q\gg 1$) и когда он значительно меньше длины волны ($q\ll 1$).

I $q\gg 1$ В этом случае наше решение по существу должно давать те же результаты, что дифракционная теория Гюйгенса — Кирхгофа или даже (при $q\to\infty$) геометрическая оптика.

Если ограничиться порядками l, значительно меньшими q и $|\hat{n}q|$, то можно использовать асимптотические приближения (73) — (76). Имеем

$$\frac{\psi_l(\hat{n}q)}{\psi_l'(\hat{n}q)}\approx\frac{i^{l+1}\exp(-i\hat{n}q)+(-i)^{l+1}\exp(i\hat{n}q)}{i^l\exp(-i\hat{n}q)+(-i)^l\exp(i\hat{n}q)}=$$

$$=\frac{\cos\left[\hat{n}q-(l+1)\frac{\pi}{2}\right]}{\cos\left[\hat{n}q-l\frac{\pi}{2}\right]}=\operatorname{tg}\left[\hat{n}q-l\frac{\pi}{2}\right]; \qquad (80)$$

I

II

III

IV

а) *б)*

Рис 13 9 Силовые линии первой (*I*), второй (*II*), третьей (*III*) и четвертой (*IV*) электрических парциальных волн [19] *а* — магнитные силовые линии, *б* — электрические силовые линии

коэффициенты (62) запишутся в виде

$$\left.\begin{aligned}
{}^eB_l&=(-1)^{l+1}\frac{2l+1}{l(l+1)}\times\\
&\times\exp(-iq)\,\frac{\sin\left[q-l\frac{\pi}{2}\right]-\hat{n}\cos\left[q-l\frac{\pi}{2}\right]\operatorname{tg}\left[\hat{n}q-l\frac{\pi}{2}\right]}{1-i\hat{n}\operatorname{tg}\left[\hat{n}q-l\frac{\pi}{2}\right]},\\
{}^mB_l&=(-1)^{l+1}\frac{2l+1}{l(l+1)}\times\\
&\times\exp(-iq)\,\frac{\hat{n}\sin\left[q-l\frac{\pi}{2}\right]-\cos\left[q-l\frac{\pi}{2}\right]\operatorname{tg}\left[\hat{n}q-l\frac{\pi}{2}\right]}{\hat{n}-i\operatorname{tg}\left[\hat{n}q-l\frac{\pi}{2}\right]}.
\end{aligned}\right\} \qquad (81)$$

Отметим, что эти коэффициенты являются быстро осциллирующими функциями и q и l, так что небольшое изменение в q или l может вызвать сильное изменение в ${}^eB_l$ и ${}^mB_l$. Видно также, что ${}^mB_{l+1}$ и ${}^eB_l$ — величины одного порядка, т. е. амплитуда электрической парциальной волны с номером l имеет тот же порядок величины, что и амплитуда магнитной парциальной волны следующего, более высокого номера.

Формулы (81) получены в предположении, что l значительно меньше q, и поэтому в таком приближении нельзя определить число парциальных волн, вносящих заметный вклад в рассеянное поле. Дебай [29] получил асимптотические приближения, справедливые для всех порядков, и показал, что амплитуды парциальных волн быстро уменьшаются до нуля как только $l+1/2$ превзойдет q, так что следует оставлять только первые q членов.

Если диэлектрическая проницаемость или проводимость сферы очень велики ($|\hat{n}|\to\infty$), выражения для ${}^eB_l$ и ${}^mB_l$ принимают вид

$$\left.\begin{aligned} {}^eB_l &= i(-1)^{l+1}\frac{2l+1}{l(l+1)}\exp(-iq)\cos\left(q-l\frac{\pi}{2}\right),\\ {}^mB_l &= (-1)^{l+1}\frac{2l+1}{l(l+1)}\exp(-iq)\sin\left(q-l\frac{\pi}{2}\right). \end{aligned}\right\} \tag{82}$$

Их проще всего получить из (63) с помощью асимптотических приближений (73) — (76).

Результаты настоящего раздела ($q\gg1$) можно применить в теории радуги [30, 31]; параметр q, определяемый размером дождевых капель, порядка 10^4.

II. $q\ll1$. Этот случай имеет большое практическое значение при изучении микроскопических и субмикроскопических частиц в коллоидных растворах. Мы можем теперь использовать разложения в степенные ряды (68) — (72) для цилиндрических функций. Ограничиваясь наиболее существенными членами, получим

$$\left.\begin{aligned} {}^eB_l &\approx i^l\frac{q^{2l+1}}{l^2[1\cdot3\cdot5\ldots(2l-1)]^2}\,\frac{\hat{n}^2-1}{\hat{n}^2+\frac{l+1}{l}},\\ {}^mB_l &\approx i^l\frac{q^{2l+3}}{l(l+1)(2l+1)(2l+3)[1\cdot3\cdot5\ldots(2l-1)]^2}(\hat{n}^2-1). \end{aligned}\right\} \tag{83}$$

Если проводимость или диэлектрическая проницаемость очень велики, из (63) найдем *)

$$\left.\begin{aligned} {}^eB_l &= i^l\frac{q^{2l+1}}{l^2[1\cdot3\cdot5\ldots(2l-1)]^2},\\ {}^mB_l &= -i^l\frac{q^{2l+1}}{l(l+1)[1\cdot3\cdot5\ldots(2l-1)]^2}. \end{aligned}\right\} \tag{84}$$

До тех пор, пока проводимость остается конечной (формулы (83)), коэффициенты ${}^eB_{l+1}$ и ${}^mB_l$ пропорциональны величине q в одной и той же степени, т. е. амплитуды $(l+1)$-й электрической и l-й магнитной парциальных волн, как и в случае, когда $q\gg1$,— величины одного порядка. Если радиус сферы так мал, что величиной q^2 можно пренебречь по сравнению с единицей, необходимо учитывать лишь первую электрическую парциальную волну; ее амплитуда и фаза определяются комплексной амплитудой

$${}^eB_1 = iq^3\frac{\hat{n}^2-1}{\hat{n}^2+2} = i\left(\frac{2\pi a}{\lambda^{(1)}}\right)^3\frac{\hat{n}^2-1}{\hat{n}^2+2}. \tag{85}$$

Поскольку ${}^eB_l$ — комплексная величина, между падающим первичным полем и рассеянным вторичным полем существует разность фаз.

*) Прямой переход от (83) к (84) не прост; необходимо принять во внимание, что когда $|\hat{n}|\to\infty$, то $q\to0$ таким образом, что $q^2|\hat{n}|^2$ стремится к конечному значению.

В другом предельном случае ($|\hat{n}|\to\infty$), как мы видим из (84), амплитуды l-й электрической и l-й магнитной парциальных волн — величины одного порядка.

Рассмотрим теперь поле первой электрической парциальной волны на большом удалении от сферы ($r \gg \lambda$) при конечном $|\hat{n}|$. Воспользуемся асимптотическим приближением

$$\zeta_1^{(1)}(x) = -\exp(ix), \qquad \zeta_1^{(1)\prime}(x) = -i\exp(ix) \tag{86}$$

и соотношениями

$$P_1(\cos\theta) = \cos\theta, \qquad P_1^{(1)}(\cos\theta) = \sin\theta, \qquad P_1^{(1)\prime}(\cos\theta)\sin\theta = -\cos\theta, \tag{87}$$

где штрих означает дифференцирование по $\cos\theta$; тогда из (58) получим

$$\left.\begin{aligned}
E_\theta^{(s)} &= -\frac{i}{k^{(\mathrm{I})}}\cos\varphi\cos\theta\,{}^eB_1\frac{\exp(ik^{(\mathrm{I})}r)}{r} = \left(\frac{2\pi}{\lambda^{(\mathrm{I})}}\right)^2 a^3\frac{\hat{n}^2-1}{\hat{n}^2+2}\cos\varphi\cos\theta\frac{\exp(ik^{(\mathrm{I})}r)}{r},\\
E_\varphi^{(s)} &= \frac{i}{k^{(\mathrm{I})}}\sin\varphi\,{}^eB_1\frac{\exp(ik^{(\mathrm{I})}r)}{r} = -\left(\frac{2\pi}{\lambda^{(\mathrm{I})}}\right)^2 a^3\frac{\hat{n}^2-1}{\hat{n}^2+2}\sin\varphi\frac{\exp(ik^{(\mathrm{I})}r)}{r},\\
H_\theta^{(s)} &= \frac{1}{k_2^{(\mathrm{I})}}\sin\varphi\,{}^eB_1\frac{\exp(ik^{(\mathrm{I})}r)}{r} = \left(\frac{2\pi}{\lambda^{(\mathrm{I})}}\right)^2\sqrt{\varepsilon^{(\mathrm{I})}}\,a^3\frac{\hat{n}^2-1}{\hat{n}^2+2}\sin\varphi\frac{\exp(ik^{(\mathrm{I})}r)}{r},\\
H_\varphi^{(s)} &= \frac{1}{k_2^{(\mathrm{I})}}\cos\varphi\cos\theta\,{}^eB_1\frac{\exp(ik^{(\mathrm{I})}r)}{r} = \left(\frac{2\pi}{\lambda^{(\mathrm{I})}}\right)^2\sqrt{\varepsilon^{(\mathrm{I})}}\,a^3\frac{\hat{n}^2-1}{\hat{n}^2+2}\cos\varphi\cos\theta\frac{\exp(ik^{(\mathrm{I})}r)}{r}.
\end{aligned}\right\} \tag{88}$$

Легко показать, что уравнения (88) идентичны уравнениям для волновой зоны ($r \gg \lambda$) *электрического диполя в точке* O с моментом $p = p_0\exp(-i\omega t)$, где

$$p_0 = a^3\left|\frac{\hat{n}^2-1}{\hat{n}^2+2}\right|, \tag{89}$$

колеблющегося параллельно оси x, т. е. параллельно электрическому вектору первичного падающего поля. Чтобы показать это, предположим вначале, что внешняя среда — вакуум ($\varepsilon^{(\mathrm{I})}=1$), и возвратимся к уравнениям (2.2.53) и (2.2.54). Согласно этим уравнениям радиационное поле линейного электрического диполя с моментом $\mathbf{p}_0\exp(-i\omega t)$ в вакууме определяется выражениями

$$\mathbf{E} = -\left(\frac{\omega}{c}\right)^2\frac{1}{r^3}\mathbf{r}\times(\mathbf{r}\times\mathbf{p}_0), \qquad \mathbf{H} = \left(\frac{\omega}{c}\right)^2\frac{1}{r^2}(\mathbf{r}\times\mathbf{p}_0). \tag{90}$$

Здесь $\mathbf{r}$ — радиус-вектор точки наблюдения. Множитель, зависящий от времени ($\exp(-i\omega t)$), опущен. Из (90) следует, что если направление вектора $\mathbf{p}_0$ совпадает с направлением оси x, то компоненты векторов $\mathbf{E}$ и $\mathbf{H}$ имеют вид

$$\left.\begin{aligned}
E_x &= \left(\frac{\omega}{c}\right)^2 p_0(\sin^2\theta\sin^2\varphi + \cos^2\theta)\frac{\exp(ikr)}{r},\\
E_y &= -\left(\frac{\omega}{c}\right)^2 p_0\sin^2\theta\sin\varphi\cos\varphi\frac{\exp(ikr)}{r},\\
E_z &= -\left(\frac{\omega}{c}\right)^2 p_0\sin\theta\cos\theta\cos\varphi\frac{\exp(ikr)}{r},\\
H_x &= 0, \qquad H_y = \left(\frac{\omega}{c}\right)^2 p_0\cos\theta\frac{\exp(ikr)}{r},\\
H_z &= -\left(\frac{\omega}{c}\right)^2 p_0\sin\theta\sin\varphi\frac{\exp(ikr)}{r}.
\end{aligned}\right\} \tag{91}$$

Если преобразовать эти выражения к сферическим координатам в соответствии с правилом (8) и подставить для p_0 его значение (89), то мы получим соотношения (88) (с $\varepsilon^{(\mathrm{I})}=1$). В более общем случае, когда $\varepsilon^{(\mathrm{I})}\neq 1$, вместо (91) необходимо рассматривать соответствующие выражения для радиационного поля диполя в диэлектрике. Эти выражения, которые можно получить так же, как и выражения (91), также преобразуются в (88) при переходе к сферическим

координатам. Отметим, что показатель преломления сферы входит в выражение (89) для эквивалентного момента диполя лишь в комбинации $(\hat{n}^2-1)/(\hat{n}^2+2)$. Для диэлектриков ($n$ вещественно) мы уже встречались с этим выражением в теории молекулярной рефракции (см. (2.3.17)).

Согласно (88) амплитуда обратно пропорциональна квадрату длины волны, и поэтому интенсивность рассеянного света обратно пропорциональна ее четвертой степени. В таком случае мы говорим о рэлеевском рассеянии *).

Подобным же образом через колебания магнитного диполя можно выразить первую магнитную парциальную волну. Парциальные волны более высоких порядков можно считать результатом колебаний мультиполей, но здесь мы не будем останавливаться на этом.

в. Интенсивность и поляризация рассеянного света. Вновь возвратимся к общему случаю и кратко исследуем интенсивность и поляризацию рассеянного света. Так как нас интересуют лишь относительные значения интенсивностей, мы вправе взять в качестве меры интенсивности квадрат вещественной амплитуды электрического вектора. Мы будем рассматривать поле лишь на больших расстояниях от рассеивающего центра ($r \gg \lambda$) и поэтому подставим в (58) асимптотические значения функций $\zeta_l^{(1)}$ и $\zeta_l^{(1)\prime}$. Положим также

$$\left.\begin{aligned} I_{\parallel}^{(s)} &= \frac{(\lambda^{(\mathrm{I})})^2}{4\pi^2 r^2}\left|\sum_{l=1}^{\infty}(-i)^l\left({}^eB_l P_l^{(1)\prime}(\cos\theta)\sin\theta - {}^mB_l\frac{P_l^{(1)}(\cos\theta)}{\sin\theta}\right)\right|^2,\\ I_{\perp}^{(s)} &= \frac{(\lambda^{(\mathrm{I})})^2}{4\pi^2 r^2}\left|\sum_{l=1}^{\infty}(-i)^l\left({}^eB_l\frac{P_l^{(1)}(\cos\theta)}{\sin\theta} - {}^mB_l P_l^{(1)\prime}(\cos\theta)\sin\theta\right)\right|^2. \end{aligned}\right\} \tag{92}$$

Тогда

$$|E_\theta^{(s)}|^2 = I_{\parallel}^{(s)}\cos^2\varphi, \qquad |E_\varphi^{(s)}|^2 = I_{\perp}^{(s)}\sin^2\varphi. \tag{93}$$

Определим *плоскость наблюдения* как плоскость, в которой лежат направление распространения падающего света и направление (θ, φ) наблюдения. Согласно (4) и (7) φ представляет собой угол между этой плоскостью и направлением колебаний электрического вектора падающей волны. На основании (93) либо $E_\theta^{(s)}$, либо $E_\varphi^{(s)}$ равно нулю, когда $\varphi = 0$ или $\varphi = \pi/2$; поэтому рассеянный свет линейно поляризован, если плоскость наблюдения параллельна или перпендикулярна первичным колебаниям. Поскольку отношение $E_\theta^{(s)}/E_\varphi^{(s)}$ комплексно, в любом другом направлении (θ, φ) свет в общем случае поляризован эллиптически. Однако в частном случае рэлеевского рассеяния (см. (88)) отношение $E_\theta^{(s)}/E_\varphi^{(s)}$ всегда вещественно и, следовательно, рассеянный свет линейно поляризован при всех направлениях наблюдения.

На практике мы обычно имеем дело с рассеянием *естественного света*. Соответствующие формулы для этого случая, как и в § 1.5, можно получить из формул для поляризованного света с помощью усреднения по всем направлениям поляризации **). Обозначая усреднение чертой и учитывая, что $\overline{\cos^2\varphi} = \overline{\sin^2\varphi} = 1/2$, мы получим вместо (93) соотношения

$$\overline{|E_\theta^{(s)}|^2} = \frac{1}{2}I_{\parallel}^{(s)}, \qquad \overline{|E_\varphi^{(s)}|^2} = \frac{1}{2}I_{\perp}^{(s)}. \tag{94}$$

*) Как мы уже упоминали на стр. 106, в своих знаменитых статьях о голубом цвете неба Рэлей показал, что спонтанные флуктуации плотности однородной среды вызывают рассеяние такого же типа [32].

**) Можно также воспользоваться результатом п. 10.8.2, согласно которому волну естественного света мы вправе считать состоящей из двух некогерентных волн равных амплитуд, распространяющихся в одном направлении и поляризованных во взаимно перпендикулярных направлениях. Рассматривая рассеяние каждой такой волны отдельно и суммируя их интенсивности, для того чтобы получить полную интенсивность, вновь придем к выражениям (94).

В общем случае ни $I_{\parallel}^{(s)}$, ни $I_{\perp}^{(s)}$ не равны нулю, так что рассеянный свет частично поляризован. По аналогии с (1.5.42) степень поляризации P рассеянного света можно определить как

$$P=\left|\frac{I_{\perp}^{(s)}-I_{\parallel}^{(s)}}{I_{\perp}^{(s)}+I_{\parallel}^{(s)}}\right|. \tag{95}$$

Неполяризованная доля рассеянного света тогда равна

$$\left.\begin{aligned}&(I_{\perp}^{(s)}+I_{\parallel}^{(s)})(1-P)=2I_{\parallel}^{(s)}, \quad \text{когда} \quad I_{\parallel}^{(s)}<I_{\perp}^{(s)},\\&(I_{\perp}^{(s)}+I_{\parallel}^{(s)})(1-P)=2I_{\perp}^{(s)}, \quad \text{когда} \quad I_{\parallel}^{(s)}>I_{\perp}^{(s)}.\end{aligned}\right\} \tag{96}$$

Зависимость интенсивности и поляризации рассеянного света от направления рассеяния и от физических параметров (λ, a, $\hat{n}$) исследовали на основе

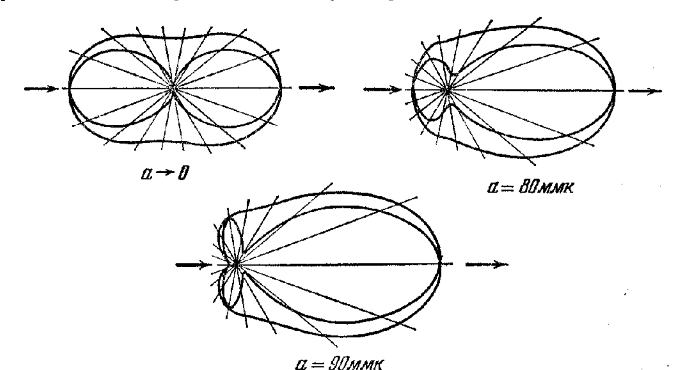


Рис. 13.10. Полярные диаграммы для рассеяния линейно поляризованного света сферической частицей золота [45].

$\lambda_0=5500$ Å, $n^{(I)}=1{,}33$, $\hat{n}^{(II)}=0{,}57+2{,}45\,i$.

теории Ми многие авторы *), однако здесь мы можем лишь кратко суммировать некоторые основные результаты.

На рис. 13.10 и 13.11 показана зависимость интенсивности, а также неполяризованной доли рассеянного света от угла наблюдения θ для диэлектрической и металлической сфер различных размеров. Длина радиуса-вектора для внешних кривых пропорциональна $I^{(s)}=I_{\parallel}^{(s)}+I_{\perp}^{(s)}$, а для внутренних кривых, если не оговорено особо, $I_{\perp}^{(s)}$. Единицы измерения произвольны и различны для каждого рисунка. Анализ этих *полярных диаграмм*, а также других опубликованных данных позволяет сделать следующие общие выводы.

За исключением случая очень большой проводимости или диэлектрической проницаемости (при этом бо́льшая часть падающего света излучается в обратном направлении, т. е. «отражается», полярные диаграммы в предельном случае исчезающе малых сфер ($a\to 0$) симметричны относительно плоскости, проходящей через центр сферы и перпендикулярной к направлению распространения падающего света. Интенсивность рассеянного света достигает максимума как в направлении, совпадающем с направлением падающего света ($\theta=0°$), так и в обратном направлении ($\theta=180°$) и имеет минимум в плоскости сим-

) В дополнение к уже цитированной работе Ми можно указать на работы [33—40]. Очень полный обзор результатов, полученных рядом исследователей, изложен в статье [22] и в книге [41]. См. также [42—44]. (На русском языке результаты, относящиеся к молекулярному рассеянию, изложены в монографии Фабелинского [83]. (*Прим. перев.*))

метрии ($\theta = 90°$). При увеличении радиуса сферы наблюдаются отклонения от симметрии, причем в направлении падения рассеивается больше света, чем в обратном направлении. Это явление часто называют *эффектом Ми*. При дальнейшем увеличении радиуса практически весь рассеянный свет будет распространяться в направлении, близком к $\theta = 0$; для проводящей сферы наибольшая концентрация света происходит также в этом направлении. Однако

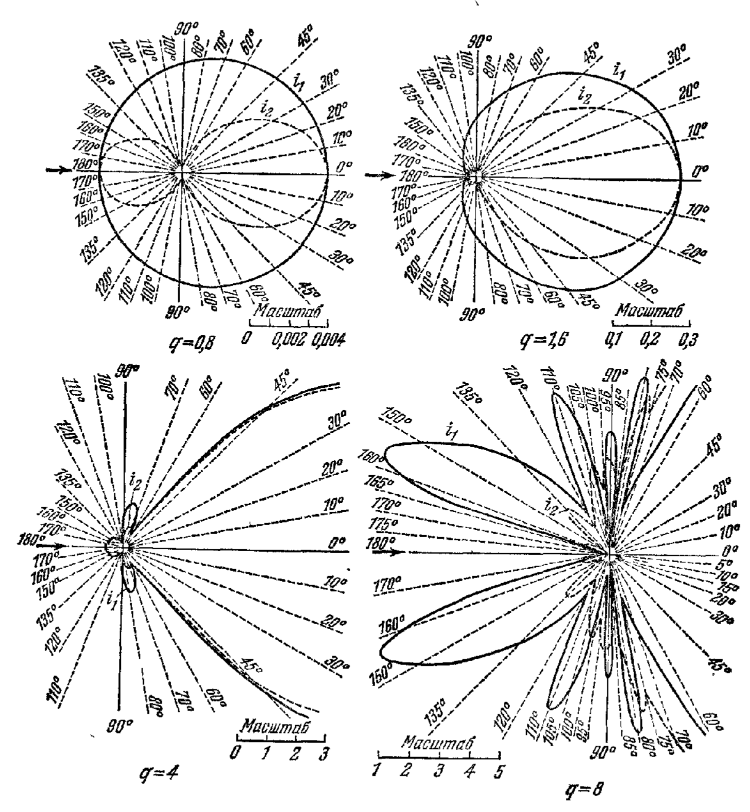


Рис. 13.11. Полярные диаграммы для рассеяния линейно поляризованного света диэлектрической сферой с показателем преломления $n = 1{,}25$ [35].
$i_1 = q^2 I_{\perp}$, $i_2 = q^2 I_{\parallel}$.

если радиус сферы очень велик по сравнению с длиной волны, то, как следует из геометрической оптики, большая часть падающего света отражается.

Зависимость интенсивности рассеянного света от радиуса сферы иллюстрирует табл. 13.4. Наличие эффекта Ми ясно видно из сравнения величин, приведенных в первом и третьем рядах. Как мы видим, имеется очень быстрый рост интенсивности с увеличением размеров сферы; чтобы получить истинные значения интенсивности, данные в таблице нужно умножить на $\lambda^{(1)2}/4\pi^2 a^2 = 1/q^2$.

Когда q превышает единицу, т. е. диаметр сферы $2a$ больше $\lambda^{(1)}/\pi$, появляется серия максимумов и минимумов, которые на первый взгляд распределе-

Таблица 13.4

Зависимость нормированных значений интенсивности $4\pi^2 a^2 (I_{\perp}^{(s)} + I_{\parallel}^{(s)})/\lambda^{(I)2}$ света, рассеянного диэлектрическими сферами с показателем преломления $n = 1{,}25$, от параметра $q = 2\pi a/\lambda^{(I)}$ [36]

θ	$q=0{,}01$	$q=0{,}1$	$q=0{,}5$	$q=1$	$q=2$	$q=5$	$q=8$
0°	$5{,}0\cdot10^{-14}$	$5{,}0\cdot10^{-8}$	$1{,}2\cdot10^{-3}$	$2{,}3\cdot10^{-1}$	4,3	$9{,}8\cdot10^{2}$	$7{,}5\cdot10^{3}$
90°	$2{,}5\cdot10^{-14}$	$2{,}5\cdot10^{-8}$	$5{,}0\cdot10^{-4}$	$3{,}6\cdot10^{-2}$	$2{,}5\cdot10^{-1}$	2,7	7,1
180°	$5{,}0\cdot10^{-14}$	$4{,}9\cdot10^{-8}$	$7{,}8\cdot10^{-4}$	$1{,}9\cdot10^{-3}$	$2{,}0\cdot10^{-2}$	1,3	0,9

ны не регулярно. Появление ряда максимумов и минимумов при больших q согласуется с теорией Гюйгенса — Кирхгофа.

Результаты, относящиеся к поляризации рассеянного света, снова оказываются различными и зависят от того, велико ли значение $|\hat{n}|$.

Для очень малых сфер с очень большой проводимостью ($\sigma\to\infty$) или очень большой диэлектрической проницаемостью ($\varepsilon\to\infty$) поляризация максимальна, когда $\theta = 60°$ (угол Томсона). С увеличением радиуса сферы ее максимум смещается в направлении увеличения θ.

Зависимость поляризации от угла наблюдения θ для сфер с конечной проводимостью и конечной диэлектрической проницаемостью показана для двух типичных случаев на рис. 13.10 и 13.11. Когда радиус сферы очень мал ($q\to 0$), диаграмма поляризации, как и диаграмма интенсивности, симметрична относительно плоскости xy и имеет максимум при $\theta = 90°$, где поляризация полная. В этом случае (рэлеевское рассеяние) степень поляризации можно записать в виде одного аналитического выражения, получающегося при подстановке (88) и (93) в (95), а именно

$$P(\theta) = \frac{\sin^2\theta}{1+\cos^2\theta}. \tag{97}$$

Эта формула была получена Рэлеем другим способом.

При увеличении радиуса сферы примерно до значения $a = \lambda^{(I)}/\pi$ максимум поляризации смещается. В большинстве исследованных случаев смещение происходит в направлении увеличения θ для диэлектрических сфер и в направлении уменьшения θ для поглощающих сфер. При дальнейшем увеличении радиуса сферы появляется нерегулярная последовательность поляризационных максимумов.

В направлении $\theta = 90°$ свет при $q < 1$ почти полностью поляризован, причем его электрический вектор перпендикулярен к плоскости наблюдения. При бóльших значениях q это уже не так, и картина становится сложнее.

До сих пор мы ограничивались случаем монохроматического света. Однако мы часто сталкиваемся с рассеянием полихроматического света, и поэтому необходимо также рассмотреть эффекты, возникающие из-за наличия компонент с различными длинами волн. Заметим, что длина волны входит в наши формулы лишь через параметр q и показатель преломления $\hat{n}$. В достаточно малой области длин волн $\hat{n}$ практически не зависит от λ, если в выражении (60) член, содержащий проводимость σ, мал по сравнению со вторым членом, т. е. в случае слабо проводящей сферы. Вместе с тем в предельном случае бесконечно большой проводимости $\hat{n}$ вообще не входит в наши формулы, и тогда интенсивности спектральных компонент зависят лишь от $a/\lambda^{(I)}$. Таким образом, влияние изменения длины волны по существу эквивалентно влиянию, вызываемому изменением на соответствующую величину радиуса сферы. Так как

для разных длин волн максимумы поляризации оказываются при различных углах наблюдения, то при наблюдении рассеянного света через поляризующую призму видны сложные изменения цвета. Этот эффект называется *полихроизмом*. Зависимость поляризации рассеянного света от длины волны — *дисперсия поляризации* — обеспечивает очень строгую проверку изложенной выше теории *).

13.5.3. Полное рассеяние и затухание. *а. Некоторые общие соображения.* Представляет значительный практический интерес определить полное количество света, которое рассеивается или поглощается сферой. Его можно найти, вычисляя вектор Пойнтинга и интегрируя его по всем направлениям. С помощью соотношений ортогональности, которые существуют между присоединенными функциями Лежандра, интегралы можно выразить через коэффициенты ${}^eB_l$ и ${}^mB_l$. Эти расчеты довольно длинны; полностью они выполнены в работе Ми [45].

Полную потерю энергии падающей волной, т. е. сумму рассеянной и поглощенной энергии, можно определить другим способом из некоторых общих соображений, применимых к телу совершенно произвольной формы. Расчеты показывают существование тесной связи между потерей энергии и амплитудой рассеянной волны в первоначальном направлении ($\theta = 0$). Из этого результата (который мы сейчас получим) и из формулы Ми для рассеянной волны легко найти полную потерю энергии на рассеяние и поглощение сферой.

Рассмотрим плоскую монохроматическую волну, падающую на небольшое тело произвольной формы, находящееся в диэлектрической среде. Поле в любой точке среды, окружающей тело, вновь можно представить в виде суммы падающего и рассеянного полей, т. е.

$$\mathbf{E} = \mathbf{E}^{(i)} + \mathbf{E}^{(s)}, \qquad \mathbf{H} = \mathbf{H}^{(i)} + \mathbf{H}^{(s)}. \tag{98}$$

Как обычно, мы опустим временной множитель $\exp(-i\omega t)$. Усредненный по времени поток энергии определяется средним значением вектора Пойнтинга. На основании (98) и формулы (1.4 56) это среднее значение равно

$$\langle \mathbf{S} \rangle = \langle \mathbf{S}^{(i)} \rangle + \langle \mathbf{S}^{(s)} \rangle + \langle \mathbf{S}' \rangle, \tag{99}$$

где

$$\langle \mathbf{S}^{(i)} \rangle = \frac{c}{8\pi} \operatorname{Re} \langle \mathbf{E}^{(i)} \times (\mathbf{H}^{(i)})^* \rangle, \tag{100а}$$

$$\langle \mathbf{S}^{(s)} \rangle = \frac{c}{8\pi} \operatorname{Re} \langle \mathbf{E}^{(s)} \times (\mathbf{H}^{(s)})^* \rangle, \tag{100б}$$

$$\langle \mathbf{S}' \rangle = \frac{c}{8\pi} \operatorname{Re} \langle \mathbf{E}^{(i)} \times (\mathbf{H}^{(s)})^* + \mathbf{E}^{(s)} \times (\mathbf{H}^{(i)})^* \rangle. \tag{100в}$$

Рассмотрим усредненный поток энергии, выходящий через поверхность сферы большого радиуса R с центром в некоторой точке области, занятой телом. Полный поток в единицу времени равен интегралу от радиальной компоненты $\langle S \rangle_r$ величины $\langle S \rangle$, взятому по сфере. Очевидно, что для диэлектрического тела он равен нулю. Однако для проводящего тела, поглощающего часть падающей энергии, суммарный поток через поверхность сферы совпадает по величине со скоростью, с которой происходит поглощение. Пусть $\mathscr{W}^{(a)}$ — скорость поглощения энергии телом. Тогда из (99) имеем

$$-\mathscr{W}^{(a)} = \mathscr{W}^{(i)} + \mathscr{W}^{(s)} + \mathscr{W}', \tag{101}$$

где $\mathscr{W}^{(i)}$, $\mathscr{W}^{(s)}$ и $\mathscr{W}'$ — интегралы от радиальных компонент $\langle \mathbf{S}^{(i)} \rangle_r$, $\langle \mathbf{S}^{(s)} \rangle_r$ и $\langle \mathbf{S}' \rangle_r$, взятые по поверхности сферы. Поскольку предполагается, что среда, окружающая тело, является непроводящей, $\mathscr{W}^{(i)} = 0$ и, следовательно,

$$\mathscr{W}^{(a)} + \mathscr{W}^{(s)} = -\mathscr{W}' = -\frac{c}{8\pi} \operatorname{Re} \iint_S (\mathbf{E}^{(i)} \times (\mathbf{H}^{(s)})^* + \mathbf{E}^{(s)} \times (\mathbf{H}^{(i)})^*) \cdot \mathbf{n}\, dS, \tag{102}$$

*) Дисперсия поляризации и полихроизм исследовались в работе [46].

где S — поверхность сферы и $\mathbf{n}$ — единичный вектор внешней нормали. Таким образом, выражение, стоящее в правой части (102), представляет скорость потери энергии на тепло и на рассеяние.

Введем единичный вектор $\mathbf{n}_0$, направление которого совпадает с направлением распространения падающей волны. Тогда

$$\mathbf{E}^{(i)} = \mathbf{e}\exp(ik^{(1)}(\mathbf{n}_0\cdot\mathbf{r})), \qquad \mathbf{H}^{(i)} = \mathbf{h}\exp(ik^{(1)}(\mathbf{n}_0\cdot\mathbf{r})). \tag{103}$$

Предположим, что эта волна линейно поляризована, и поэтому $\mathbf{e}$ и $\mathbf{h}$ можно считать вещественными постоянными векторами. На больших расстояниях от тела рассеянную волну можно считать сферической, т. е.

$$\mathbf{E}^{(s)} = \mathbf{a}(\mathbf{n})\frac{\exp(ik^{(1)}r)}{r}, \qquad \mathbf{H}^{(s)} = \mathbf{b}(\mathbf{n})\frac{\exp(ik^{(1)}r)}{r}. \tag{104}$$

Векторы $\mathbf{a}(\mathbf{n})$ и $\mathbf{b}(\mathbf{n})$ характеризуют силу излучения, рассеянного в направлении $\mathbf{n}$. Так как и падающая, и рассеянная волны подчиняются уравнениям Максвелла, получим (см. уравнения (1.4.4) и (1.4.5))

$$\left.\begin{aligned} \mathbf{h} &= \sqrt{\varepsilon^{(1)}}\,\mathbf{n}_0\times\mathbf{e}, & \mathbf{b} &= \sqrt{\varepsilon^{(1)}}\,\mathbf{n}\times\mathbf{a},\\ \mathbf{n}_0\cdot\mathbf{e} &= \mathbf{n}_0\cdot\mathbf{h} = 0, & \mathbf{n}\cdot\mathbf{a} &= \mathbf{n}\cdot\mathbf{b} = 0, \end{aligned}\right\} \tag{105}$$

где $\varepsilon^{(1)}$ — диэлектрическая проницаемость внешней среды, которая предполагается немагнитной ($\mu = 1$). Из этих соотношений следует, что на поверхности S сферы большого радиуса

$$\left.\begin{aligned} &(\mathbf{E}^{(i)}\times(\mathbf{H}^{(s)})^*)\cdot\mathbf{n} = \sqrt{\varepsilon^{(1)}}\,\mathbf{e}\cdot\mathbf{a}^*\exp[ik^{(1)}R(\mathbf{n}_0\cdot\mathbf{n})]\frac{\exp[-ik^{(1)}R]}{R},\\ &(\mathbf{E}^{(s)}\times(\mathbf{H}^{(i)})^*)\cdot\mathbf{n} =\\ &= \sqrt{\varepsilon^{(1)}}\,[(\mathbf{n}\cdot\mathbf{n}_0)(\mathbf{a}\cdot\mathbf{e}) - (\mathbf{n}\cdot\mathbf{e})(\mathbf{n}_0\cdot\mathbf{a})]\exp[-ik^{(1)}R(\mathbf{n}_0\cdot\mathbf{n})]\frac{\exp[ik^{(1)}R]}{R}. \end{aligned}\right\} \tag{106}$$

Подставим найденные выражения в (102). Для вычисления полученного интеграла воспользуемся леммой *), которая утверждает, что для произвольной функции $f(\mathbf{n})$ при больших R можно написать

$$\frac{1}{R}\iint_S f(\mathbf{n})\exp[-ik^{(1)}R(\mathbf{n}_0\cdot\mathbf{n})]\,dS \approx \\ \approx \frac{2\pi i}{k^{(1)}}[f(\mathbf{n}_0)\exp(-ik^{(1)}R) - f(-\mathbf{n}_0)\exp(ik^{(1)}R)]. \tag{107}$$

Тогда мы получим

$$\left.\begin{aligned} &\iint_S (\mathbf{E}^{(i)}\times(\mathbf{H}^{(s)})^*)\cdot\mathbf{n}\,dS \approx\\ &\qquad \approx -\frac{2\pi i}{k^{(1)}}\sqrt{\varepsilon^{(1)}}\,[\mathbf{e}\cdot\mathbf{a}^*(\mathbf{n}_0) - \mathbf{e}\cdot\mathbf{a}^*(-\mathbf{n}_0)\exp(-2ik^{(1)}R)],\\ &\iint_S (\mathbf{E}^{(s)}\times(\mathbf{H}^{(i)})^*)\cdot\mathbf{n}\,dS \approx \frac{2\pi i}{k^{(1)}}\sqrt{\varepsilon^{(1)}}\,[\mathbf{e}\cdot\mathbf{a}(\mathbf{n}_0) + \mathbf{e}\cdot\mathbf{a}(-\mathbf{n}_0)\exp(2ik^{(1)}R)], \end{aligned}\right\} \tag{108}$$

а соотношение (102) примет вид

$$\mathscr{W}^{(s)} + \mathscr{W}^{(a)} = \frac{c\sqrt{\varepsilon^{(1)}}}{2k^{(1)}}\operatorname{Im}[\mathbf{e}\cdot\mathbf{a}(\mathbf{n}_0)], \tag{109}$$

где Im означает, что берется мнимая часть.

Из соотношения (109) следует, что *в случае падения линейно поляризованного света скорость диссипации энергии пропорциональна проекции на направление электрического вектора падающей волны амплитуды волны, рассеянной в первоначальном направлении* ($\mathbf{n} = \mathbf{n}_0$).

*) Эту лемму можно доказать после очевидной замены переменных с помощью принципа стационарной фазы (см. приложение 3). См. также [47].

Величина Q, равная отношению скорости диссипации энергии $(\mathscr{W}^{(a)}+\mathscr{W}^{(s)})$ к количеству энергии, падающей в единицу времени на единичную площадку в сечении тела $(|\langle \mathbf{S}^{(i)}\rangle|)$, называется *сечением экстинкции* тела. Из (100а), (103) и (105) следует, что $|\langle \mathbf{S}^{(i)}\rangle| = c\sqrt{\varepsilon^{(\mathrm{I})}}\mathbf{e}^2/8\pi$, так что, согласно (109), найдем

$$Q = \frac{\mathscr{W}^{(s)}+\mathscr{W}^{(a)}}{|\langle \mathbf{S}^{(i)}\rangle|} = 2\lambda^{(\mathrm{I})}\,\mathrm{Im}\left(\frac{\mathbf{e}\cdot\mathbf{a}(\mathbf{n}_0)}{\mathbf{e}^2}\right). \qquad (110)$$

Эта формула получена Ван Хулстом *) [48].

Аналогичным образом можно определить *сечение рассеяния* $Q^{(s)}$ и *сечение поглощения* $Q^{(a)}$ тела, а именно

$$Q^{(s)} = \frac{\mathscr{W}^{(s)}}{|\langle \mathbf{S}^{(i)}\rangle|}, \qquad Q^{(a)} = \frac{\mathscr{W}^{(a)}}{|\langle \mathbf{S}^{(i)}\rangle|}. \qquad (111)$$

Очевидно, что $Q = Q^{(s)} + Q^{(a)}$. Для непоглощающего тела $Q^{(a)} = 0$, и сечение экстинкции совпадает с сечением рассеяния.

Прежде чем применить соотношение (110) к сферическому телу, найдем величину Q для тела, которое практически не пропускает падающего света. Предположим также, что его линейные размеры велики по сравнению с длиной волны. В этом случае применима теория Гюйгенса — Кирхгофа, и основной вклад в рассеяние в направлении падения света обусловлен дифракцией Фраунгофера. Пусть на тело падает линейно поляризованная волна с плоским волновым фронтом, причем $\mathscr{A}'$ — часть этого фронта, не закрываемая препятствием, и $\mathscr{A}$ — его часть, занятая препятствием (рис. 13.12). Рассмотрим рассеянное поле $\mathbf{E}^{(s)}$ в точке P, находящейся на большом расстоянии от тела. Согласно принципу Гюйгенса — Френеля и принципу Бабине (8.3.21) имеем для малого угла дифракции

$$\mathbf{E}^{(s)} = \frac{i}{\lambda^{(\mathrm{I})}}\mathbf{e}\iint\limits_{\mathscr{A}} \frac{\exp(ik^{(\mathrm{I})}r)}{r}\,dS. \qquad (112)$$

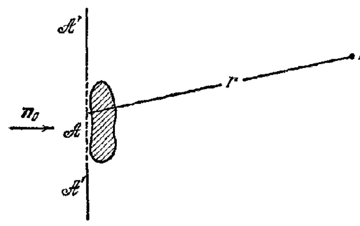


Рис. 13.12. К выводу соотношения (114) для сечения экстинкции при большом препятствии

Если точка P находится очень далеко от препятствия в направлении распространения падающей волны, то r можно считать постоянным, и (112) дает

$$\mathbf{E}^{(s)}(\mathbf{n}_0) = \frac{i}{\lambda^{(\mathrm{I})}}D\mathbf{e}\frac{\exp(ik^{(\mathrm{I})}r)}{r}; \qquad (113)$$

здесь D — геометрическое поперечное сечение тела (площадь $\mathscr{A}$). Следовательно, в данном случае вектор $\mathbf{a}$ в (104) как функция $\mathbf{n}_0$ определяется выражением $iD\mathbf{e}/\lambda^{(\mathrm{I})}$ и из (110) находим, что

$$Q = 2D. \qquad (114)$$

Таким образом, *сечение экстинкции большого непрозрачного тела равно удвоенному значению его геометрического поперечного сечения*. На первый взгляд полученный результат кажется несколько парадоксальным, так как можно было бы ожидать, что для тела большего размера должно было бы быть справедливым приближение геометрической оптики, а в этом приближении сечение экстинкции равно D. Объяснение этого кажущегося противоречия **) заключается в том, что независимо от размеров тела и расстояния до точки, в которой рассматривается поле, всегда имеется хотя бы узкая область — вблизи края геометрической тени,— где приближение геометрической оптики несправедливо.

*) Используемый здесь метод доказательства предложен в работе [49]. Аналогичная теорема существует и в теории атомных столкновений [50, 51].

**) Аналогичный кажущийся парадокс, впервые отмеченный в работе [52], возникает в проблемах рассеяния в квантовой механике.

Помимо света, задерживаемого телом с поперечным сечением D (потери на отражение и поглощение), возникает дополнительный вклад в экстинкцию, обусловленный соседством с краем тени, и очевидно, что этот вклад также равен D. Для экспериментальной проверки соотношения (114) необходимо собрать свет с достаточно широкой площадки, расположенной достаточно далеко от тела *).

Применим теперь общую формулу (110) к сферическому телу. На основании (58) проекция амплитуды рассеянной волны на направление электрического вектора падающей волны ($\varphi = 0$) при рассеянии вперед ($\theta = 0$) определяется выражением

$$(E_\theta^{(s)})_{\theta=\varphi=0} = -\frac{1}{k^{(1)}}\frac{1}{r}\sum_{l=1}^{\infty}\Big\{{}^eB_l\,\zeta_l^{(1)\prime}(k^{(1)}r)\,[P_l^{(1)\prime}(\cos\theta)\sin\theta]_{\theta=0} - i\,{}^mB_l\,\zeta_l^{(1)}(k^{(1)}r)\left[P_l^{(1)}(\cos\theta)\frac{1}{\sin\theta}\right]_{\theta=0}\Big\}. \quad (115)$$

Оба члена, содержащих присоединенные полиномы Лежандра, легко найти из выражения **) [23]

$$P_l^{(m)}(x) = \frac{1}{2^m m!}\frac{(l+m)!}{(l-m)!}(1-x^2)^{m/2}\left\{1 + c_1\left(\frac{1-x}{2}\right) + c_2\left(\frac{1-x}{2}\right)^2 + \ldots\right\}; \quad (116)$$

здесь $c_1, c_2, \ldots$ зависят лишь от l и m. Из последнего выражения и из его производной находим

$$\left(\frac{P_l^{(1)}(\cos\theta)}{\sin\theta}\right)_{\theta=0} = \frac{1}{2}l(l+1), \qquad (P_l^{(1)\prime}(\cos\theta)\sin\theta)_{\theta=0} = -\frac{1}{2}l(l+1). \quad (117)$$

Подставляя (117) в (115) и используя для $\zeta_l^{(1)}$ и $\zeta_l^{(1)\prime}$ асимптотические приближения (74) и (76), получим

$$(E_\theta^{(s)})_{\theta=\varphi=0} = \frac{1}{2k^{(1)}}\frac{\exp(ik^{(1)}r)}{r}\sum_{l=1}^{\infty}(-i)^l\,l(l+1)\,[{}^eB_l + {}^mB_l]. \quad (118)$$

Вспоминая, что амплитуда падающего поля считается равной единице ($\mathbf{e}^2 = 1$), находим искомую величину $\mathbf{e}\cdot\mathbf{a}(\mathbf{n}_0)/\mathbf{e}^2$, разделив найденное выражение на $\exp(ik^{(1)}r)/r$. Подставляя ее в (110) и используя тождество $\operatorname{Im}(z) \equiv \operatorname{Re}(-iz)$, справедливое для любого комплексного числа z, мы окончательно получим следующие выражения для *сечения экстинкции сферы*:

$$Q = \frac{\lambda^{(1)2}}{2\pi}\operatorname{Re}\sum_{l=1}^{\infty}(-i)^{l+1}\,l(l+1)\,[{}^eB_l + {}^mB_l], \quad (119)$$

здесь коэффициенты ${}^eB_l$ и ${}^mB_l$ определяются выражениями (62).

б. *Результаты расчетов.* Суммируем теперь основные результаты расчетов полного рассеяния, полного поглощения и экстинкции для сферы.

В п. 13.5.2 мы показали, что при диаметре сферы, значительно меньшем длины волны (рэлеевское рассеяние), необходимо учитывать лишь первую электрическую парциальную волну. В этом случае амплитуда рассеянной волны пропорциональна $1/(\lambda^{(1)})^2$, так что полное рассеяние обратно пропорционально четвертой степени длины волны. Если учитывать также члены более высокого порядка, которые зависят от радиуса сферы и материальных постоянных, то полное рассеяние станет очень сложной функцией длины волны и будет

*) Тщательный анализ, приводящий к лемме (107), показывает, что часть поверхности сферы S, вносящая заметный вклад в Q, стягивает центральный телесный угол порядка $(kR)^{-\beta}$, где $1 > \beta > 4/5$ (см. [47]). Более подробное исследование вопроса о сечении экстинкции большой частицы и роли геометрической тени можно найти в работах [53, 54].

**) Формула, приведенная в [23], отличается от (116) множителем $(-1)^m$. Различие связано с тем, что в работе [23] использовано несколько иное определение присоединенных полиномов Лежандра (см. сноску на стр. 597).

селективным *). В случае золота, например, даже очень маленькая сфера дает максимум вблизи $\lambda = 5500$ Å (рис. 13.13).

Такой максимум можно интерпретировать как своего рода резонанс. Предположим, что сфера не испытывает влияния поля падающего пучка света, но в ней возбуждены свободные электромагнитные колебания. Частоту и постоянную затухания таких свободных колебаний можно получить из изложенной выше теории, если в уравнениях (56) опустить члены, не содержащие коэффициентов ${}^eB_l$ и ${}^mB_l$. Получающаяся система линейных однородных уравнений допускает нетривиальное решение только в том случае, если они совместны. Тогда каждое решение уравнения соответствует затухающим собственным колебаниям; частота таких колебаний очень близка к частотам, на которых имеется максимум интенсивности для определенных рассеянных парциальных волн **).

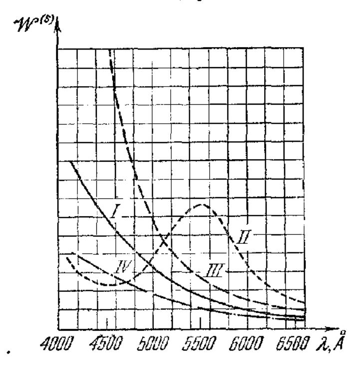


Рис. 13.13. Зависимость полного рассеяния на очень маленьких сферах ($a \to 0$) от длины волны [56].

I — ртуть, *II* — золото, *III* — серебро, *IV* — идеальный проводник.

Расчеты рассеяния и экстинкции сферами конечного радиуса для выбранных значений показателя преломления были проделаны многими авторами. Большинство расчетов относится к диэлектрическим (вещественное n) и слабо поглощающим сферам. На рис. 13.14 приведена типичная кривая, полученная Гольдбергом для сфер с показателем преломления ***) $n = 1{,}33$. Такой же показатель преломления у воды, и поэтому указанные результаты представляют интерес в связи с пропусканием света дымкой, тучами и туманом, а также

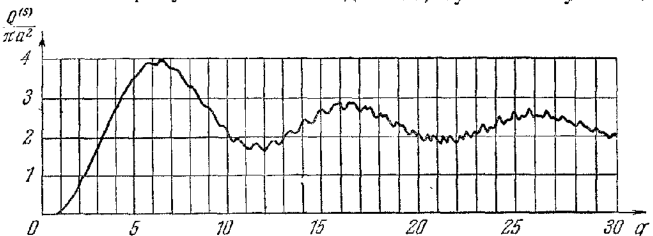


Рис. 13.14. Зависимость сечения рассеяния диэлектрических сфер с показателем преломления $n = 1{,}33$ от параметра $q = 2\pi a/\lambda^{(1)}$ [59].

в связи с теорией камеры Вильсона и т. д. Как мы видим, на кривой имеется серия максимумов и минимумов, и с увеличением радиуса сферы сечение экстинкции в согласии с уравнением (114) стремится к удвоенному значению гео-

*) Интересным примером селективного рассеяния служат явления, наблюдавшиеся в сентябре 1950 г., когда на большой части Европы Солнце (а также Луна) казалось насыщенно голубым. Спектроскопические измерения «голубого» и нормального Солнца позволили получить кривую экстинкции слоя, вызывающего это явление. Было сделано заключение, что голубой цвет обусловлен селективным рассеянием дымом, вероятно, состоящим из поразительно одинаковых по размеру капелек масла, занесенных ветрами, дующими в верхних слоях атмосферы, из горящих лесов в Альберте [55].

**) Оптический резонанс наблюдал Вуд [57] на гранулированных пленках и парах щелочных металлов. Собственные колебания сферы изучались Дебаем [58].

***) Этот случай исследовался также в работах [60, 61]. Сходные кривые, относящиеся к другим значениям показателя преломления, были получены многими авторами, например, [22, 41, 62—65].

метрического сечения. На кривой заметна также тонкая структура, т. е. дополнительные небольшие максимумы и минимумы. Естественно, что эти небольшие флуктуации сглаживаются, если рассеяние вызывается многими сферическими частицами не точно одинакового размера.

Аналогичный характер имеют кривые экстинкции для диэлектрических сфер с другими показателями преломления. Можно показать, что если n не слишком отличается от единицы, первый максимум у всех кривых получается при значении q, определяемом выражением *) $2q(n-1)\approx 4$, и Q при этом может достигать величины $4\pi a^2$.

Для полностью отражающих сфер [64, 66] ($n\to\infty$) первый максимум на кривой экстинкции оказывается при $q=1{,}2$, а $Q=2{,}29\,\pi a^2$, первый минимум — при $q=1{,}6$, а $Q=2{,}12\,\pi a^2$. Затем появляются слабые осцилляции и кривая приближается к значению $2\pi a^2$, когда $q\to\infty$.

Расчеты, относящиеся к случаю поглощающих сфер, значительно более трудоемки, и поэтому подробно изучено лишь несколько специальных случаев. На рис. 13.15 показаны кривые, относящиеся к рассеянию, поглощению и

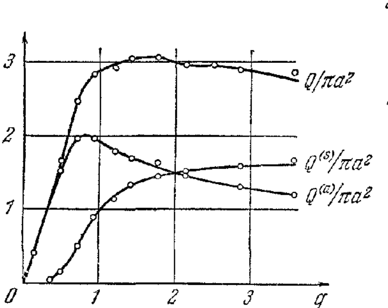


Рис. 13.15. Сечение поглощения $Q^{(a)}$, сечение рассеяния $Q^{(s)}$ и сечение экстинкции Q для железных сфер различного радиуса [67].

$\hat{n}=1{,}27+1{,}37\,i$, $\lambda^{(1)}=4200$ Å.

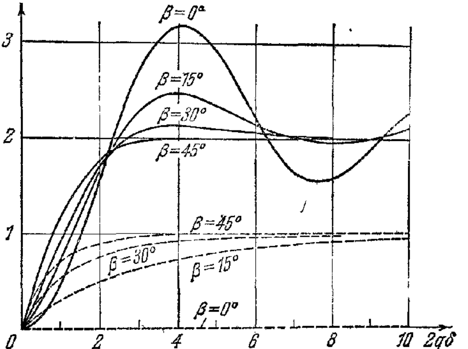


Рис. 13.16. Кривые экстинкции $Q/\pi a^2$ (сплошные линии) и кривые поглощения $Q^{(a)}/\pi a^2$ (пунктирные линии) для слабо поглощающих сфер с показателем преломления $\hat{n}=(1+\delta)+i\delta\,\mathrm{tg}\,\beta$, где величина δ вещественна и мала по сравнению с единицей [69]

экстинкции небольшими сферами железа. Для сфер с большим диаметром при расчете можно пользоваться асимптотическими формулами [68], выведенными на основе теории Ми и асимптотических разложений Дебая для цилиндрических функций. Случай слабо поглощающих сфер рассматривался Хулстом [69], результаты его расчетов показаны на рис. 13.16. Как мы видим, в последнем случае общее поведение кривых экстинкции, аналогично их поведению для диэлектрических сфер, но уже очень небольшой проводимости достаточно, чтобы полностью сгладить имеющиеся слабые осцилляции. При дальнейшем увеличении проводимости первый минимум совершенно исчезает и функции, описывающие экстинкцию и поглощение, монотонно увеличиваются, асимптотически приближаясь к значениям 2 и 1 соответственно.

Теорию Ми можно проверить экспериментально путем наблюдения рассеяния света либо одной сферической частицей, либо многими частицами (мутные среды, коллоидные растворы). Такую проверку относительно легко

*) Величина $2q(n-1)$ представляет фазовый сдвиг, который претерпевает луч света, пересекающий сферу по диаметру. Это полезный параметр, и если строить кривые экстинкции в зависимости от него, то при n, не слишком отличающемся от единицы, они чрезвычайно похожи друг на друга.

провести для больших частиц, но она становится довольно трудной, когда диаметр каждой частицы порядка длины волны или меньше. Ла Меру с сотрудниками [62, 63, 70] удалось проверить теорию, измеряя угловое распределение рассеянного света, а также полное рассеяние от взвеси серных золей в воде (частицы диаметром от 3000 до 5000 Å). Они использовали излучение с длинами волн в вакууме от 2850 до 10000 Å и нашли прекрасное согласие с предсказаниями теории Ми. В некоторых случаях на кривой экстинкции наблюдались даже мелкие флуктуации (тонкая структура) (см. рис. 13.14).

Ряд авторов изучал рассеяние света несферическими частицами, но в общем случае аналитический вид соответствующих волновых функций настолько сложен, что строгие решения имеют ограниченное практическое значение *). Ганс [75] и другие исследователи рассматривали рассеяние электромагнитных волн эллипсоидами с размерами, малыми по сравнению с длиной волны; строгое решение для эллипсоида произвольного размера было получено в работе [76]. Рассеяние длинными круглыми проводящими цилиндрами изучалось еще в 1905 г. Зейтцем [77] и Игнатовским [78], и полученные ими формулы подобны формулам Ми для сферы. Рассеяние длинными круглыми диэлектрическими цилиндрами и цилиндрами с высоким отражением исследовали Шеффер и Гроссманн [79] (см. также [80]).

ЛИТЕРАТУРА

1. A. Kundt, Ann. d. Physik 34, 469 (1888).
2. P. Drude, Ann. d. Physik 39, 504 (1890).
3. J. R. Collins, R. O. Bock, Rev. Sci. Instr. 14, 135 (1943).
4. I. Simon, J. Opt. Soc. Amer. 41, 336 (1951).
5. D. G. Avery, Proc. Phys. Soc. 65, 425 (1952).
6. R. W. Ditchburn. J. Opt. Soc. Amer. 45, 743 (1955).
7. H. H. Landolt, R. Börnstein. Phys. Chem. Tabellen, 5 Aufl., Berlin, 1923; 1—3 Ergänzungsb., 1927—1936.
8. E. Hagen, H. Rubens, Ann. d. Physik 11, 873 (1903).
9. M. Planck, Theory of Heat, McMillan, London, 1932. (М. Планк, Введение в теоретическую физику, ч. 5, Теория теплоты, ОНТИ, 1935.)
10. A. Sommerfeld, Thermodynamics and statistical Mechanics, ed. by F. Bopp a. J. Meixner, Acad. Press, N. Y., 1956. (А. Зоммерфельд, Термодинамика и статистическая физика, ИЛ, 1955.)
11. M. Born, Atomic Physics, Blackie a. Son, London, 5th ed., 1951, (М. Борн, Атомная физика, «Мир», 1965.)
12. K. Hammer, Z. Tech. Phys. 24, 169 (1943).
13. F. Abelès, Rev. d.'Opt. 32, 257 (1953).
14. L. N. Hadley. D. M. Dennison, J. Opt. Soc. Amer. 37, 451 (1947); 38, 483 (1948).
15. S. Tolansky, Multiple-beam Interferometry of Surfaces and Films, Oxford Univ. Press, 1948, p. 26.
16. O. S. Heavens, Optical Properties of Thin Solid Films, Butterworths, London, 1955.
17. H. Schopper, Optik 10, 426 (1953).
18. J. C. M. Garnett, Phil. Trans. Roy. Soc. A203, 385 (1904); 205, 237 (1906).
19. G. Mie, Ann. d. Physik 25, 377 (1908).
20. P. Debye, Ann. d. Physik. 30, 57 (1909).
21. T. J. I'A. Bromwich, Phil. Trans. Roy. Soc. A220, 175 (1920).
22. H. C. van de Hulst, Rech. Astron. Observ. Utrecht 11, Pt I (1946).
23. W. Magnus, F. Oberhettinger, Formulas and Theorems for the Functions of Mathematical Physics, Chelsea Publ. Co., N. Y., 1954.
24. P. Frank, R. Mises, Riemann — Weber's Differentialgleichungen der Mathematischer Physik, Vieweg, Braunschweig, 2nd ed., 1935. (Ф. Франк, Р. Мизес, Дифференциальные и интегральные уравнения математической физики, Гостехиздат, 1937.)
25. C. J. Bouwkamp, H. B. J. Casimir, Physica 20, 539 (1954); A. Nisbet, Proc. Roy. Soc. A231, 260 (1955); Physica 21, 799 (1955).
26. A. Sommerfeld, Partial Differential Equations of Physics, Acad. Press, N. Y., 1949. (А. Зоммерфельд, Дифференциальные уравнения в частных производных физики, ИЛ, 1950.)

*) Некоторыми авторами развивались приближенные методы (см., например, [71—73]). См. также обзорную статью [74].

27. G. N. Watson, A Treatise on the Theory of Bessel Functions, Cambr. Univ. Press, 2nd ed., 1944. (Г. Н. Ватсон, Теория бесселевых функций, ИЛ, 1949.)
28. K. Schwarzschild, Münch. Akad., Math.-Phys. Kl. 31, 293 (1901).
29. P. Debye, Ann. d. Physik 30, 118 (1909); Math. Ann. 67, 535 (1909); Sitzungsb. Münch. Akad. Wiss., Math.-Phys. Kl., 5 Abh. (1910).
30. B. van der Pol, M. Bremmer, Phil. Mag. 24, 857 (1937).
31. H. Buccrius, Optik 1, 188 (1946).
32. Rayleigh. Phil. Mag. 41, 274, 447 (1871); 47, 375 (1899); Scientific Papers 1, 87, 104 (1899); 4, 397 (1903).
33. R. Gans, Ann. d. Physik 76, 29 (1925).
34. H. Senftleben, E. Benedict, Ann. d. Physik 60, 297 (1919).
35. H. Blumer, Z. f. Phys. 32,, 119 (1925).
36. H. Blumer, Z. f. Phys. 38, 304, 920 (1926).
37. H. Blumer, Z. f. Phys, 39, 195 (1926).
38. C. Schalén, Uppsala Astr. Obs. Ann. 1, № 2 (1939); № 9 (1945).
39. G. R. Paranjpe, Y. G. Naik, P. B. Vaidya, Proc. Indian Acad. A9, 333, 352 (1939).
40. H. Holl, Optik 1, 213 (1946); 4, 173 (1948/49).
41. H. C. van de Hulst, Light Scattering by Small Particles, John Wiley a. Sons, N. Y., Chapman a. Hall, London, 1957.
42. G. Oster, Chem. Rev. 43, 319 (1948).
43. Tables of scattering functions for spherial particles, Nat. Bureau Stand. Applied Mathematics Series, 4, Washington, 1949.
44. R. O. Gumprecht, C. M. Sliepcevich, Light-Scattering Functions for Spherical Particles, Univ. of Michigan Press, Ann. Arbor. 1951.
45. G. Mie, Ann. d. Physik 25, 429 (1908).
46. M. A. Schirmann, Ann. d. Physik 59, 493 (1919).
47. D. S. Jones, Proc. Cambr. Phil. Soc 48, 736 (1952).
48. H. C. van de Hulst, Physica 15, 740 (1949).
49. D. S. Jones, Phil. Mag. 46, 957 (1955).
50. E. Feenberg, Phys. Rev. 40, 48 (1932).
51. M. Lax, Phys. Rev. 78, 306 (1950).
52. H. S. W. Massey, C. B. O. Mohr, Proc. Roy. Soc. A141, 434 (1933).
53. D. Sinclair, J. Opt. Soc. Amer. 37, 475 (1947).
54. L. Brillouin, J. Appl. Phys. 20, 1110 (1949).
55. R. Wilson, Mo. Not. Roy. Astr. Soc. 111, 478 (1951).
56. R. Feick, Ann. d. Physik 77, 582 (1925).
57. R. W. Wood, Phil. Mag. 3, 396 (1902).
58. P. Debye, Ann. d. Physik, 30, 73 (1909).
59. B. Goldberg, J. Opt. Soc. Amer. 43, 1221 (1953).
60. H. Holl, Optik 4, 173 (1948).
61. H. G. Houghton, W. R. Chalker, J. Opt. Soc. Amer. 39, 955 (1949).
62. M. D. Barnes, V. K. La Mer, J. Colloid Sci, 1, 79 (1946).
63. M. D. Barnes, A. S. Kenyon, E. M. Zaiser, V. K. La Mer, J. Colloid Science 2, 349 (1947).
64. J. L. Greenstein, Harvard circ., 1937, № 422.
65. R. Pendorf, J. Opt. Soc. Amer. 46, 1001 (1956).
66. F. W. A. Götz, Astr. Nachr. 255, 63 (1935).
67. C. Schalén, Uppsala Astr. Observ. Ann. 1, № 9 (1945).
68. G. Jobst, Ann. d. Physik, 76, 863 (1925).
69. H. C. van de Hulst, Rech. Astr. Observ. Utrecht 11, Pt. 2, 27 (1949).
70. I. Johnson, V. K. La Mer, J. Amer. Chem. Soc. 69, 1184 (1947).
71. R. W. Hart, E. W. Montroll, J. Appl. Phys. 22, 376 (1951).
72. E. W. Montroll, R. W. Hart, J. Appl. Phys. 22,, 1278 (1951).
73. E. W. Montroll, J. M. Greenberg, Phys. Rev. 86, 889 (1952).
74. C. J. Bouwkamp, в сб. «Rep. Progr. Phys». Phys. Soc. London, 1954, Vol. 17. p. 35.
75. R. Gans, Ann. d. Physik 37, 881 (1912); 47, 270 (1915).
76. F. Möglich, Ann. d. Physik 83, 609 (1927).
77. W. Seitz, Ann. d. Physik 16, 746 (1905); 19, 554 (1906).
78. W. Ignatowsky, Ann. d. Physik 18, 495 (1905).
79. C. Schaeffer, F. Grossmann, Ann. d. Physik 31, 455 (1910).
80. H. C. van de Hulst, Astrophys. J. 112, 1 (1950).
81* В. Л. Гинзбург, Г. П. Мотулевич, УФН 55, 469 (1955).
81а*. Г. П. Мотулевич, УФН 97, 211, 19697.
82*. А. В. Соколов, Оптические свойства металлов, Физматгиз, 1961.
83*. И. Л. Фабелинский. Молекулярное рассеяние света, «Наука», 1965.

ГЛАВА 14

КРИСТАЛЛООПТИКА*)

§ 14.1. Тензор диэлектрической проницаемости анизотропной среды

Вспомним, что в основе нашей оптической теории лежат две отдельные системы: одна — уравнения Максвелла (1.1.1) и (1.1.2), другая — материальные уравнения, которые для изотропной среды были записаны в виде формул (1.1.9) — (1.1.11). Чтобы принять во внимание анизотропию кристаллов, необходимо обобщить последние уравнения. В основной части настоящей главы мы будем рассматривать однородную, непроводящую ($\sigma = 0$) и магнитно изотропную среду **), считая ее, однако, *электрически анизотропной*. Иными словами, мы будем рассматривать вещества, электрическое возбуждение которых зависит от направления электрического поля. Тогда вектор **D**, вообще говоря, не будет параллелен вектору **E**. Заменим уравнение (1.1.10) простейшим соотношением между **D** и **E**, позволяющим учесть анизотропию, а именно соотношением, в котором каждая компонента вектора **D** связана линейно с компонентами **E**, т. е.

$$\left.\begin{aligned} D_x &= \varepsilon_{xx}E_x + \varepsilon_{xy}E_y + \varepsilon_{xz}E_z, \\ D_y &= \varepsilon_{yx}E_x + \varepsilon_{yy}E_y + \varepsilon_{yz}E_z, \\ D_z &= \varepsilon_{zx}E_x + \varepsilon_{zy}E_y + \varepsilon_{zz}E_z. \end{aligned}\right\} \qquad (1)$$

Девять величин ε_{xx}, ε_{yy}, ... являются постоянными среды и составляют *тензор диэлектрической проницаемости*; следовательно, вектор **D** равен произведению этого тензора на вектор **E**.

Перепишем уравнение (1) в более компактной форме:

$$D_k = \sum_l \varepsilon_{kl} E_l, \qquad (2)$$

где k — один из трех индексов x, y или z, а индекс l, по которому ведется суммирование, принимает по очереди значения x, y и z. При формальной тензорной записи знак суммирования обычно опускают, а указанием на суммирование по всем значениям l служит двукратное появление в произведении этого индекса. Однако мы сохраним знак суммы, так как это поможет избежать каких-либо неясностей для читателей, не знакомых с тензорным исчислением.

Предположим, что выражения (1.1.31) для плотностей электрической и магнитной энергий остаются справедливыми и здесь. Тогда

$$w_e = \frac{1}{8\pi}\mathbf{E}\cdot\mathbf{D} = \frac{1}{8\pi}\sum_{kl} E_k \varepsilon_{kl} E_l \qquad (3)$$

и

$$w_m = \frac{1}{8\pi}\mathbf{B}\cdot\mathbf{H} = \frac{1}{8\pi}\mu H^2. \qquad (4)$$

Сохраним также определение вектора Пойнтинга, или «лучевого вектора»,

---

) В настоящей главе рассматриваются вопросы кристаллооптики без учета пространственной дисперсии. Учет пространственной дисперсии в кристаллооптике выполнен в книге В. М. Аграновича и В. Л. Гинзбурга [40]. (*Прим. ред.*).

**) Существуют также магнитные кристаллы, но поскольку влияние намагничения на оптические явления мало (быстрые осцилляции), магнитной анизотропией можно пренебречь. Однако чтобы сохранить некоторую симметрию формул и включить слабо магнитные кристаллы, мы учтем магнитную проницаемость, считая ее скалярной величиной μ; кроме того, это облегчит написание уравнений в тех системах единиц, где μ для вакуума не равно единице.

задаваемое (1.1.38), т. е.

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E}\times\mathbf{H}, \tag{5}$$

и посмотрим, согласуются ли эти определения с законом сохранения энергии.

Как и в п. 1.1.4, умножим первое уравнение Максвелла на **E**, второе на **H** и используем векторное тождество (1.1.27). Тогда получим

$$-c\operatorname{div}(\mathbf{E}\times\mathbf{H}) = \mathbf{E}\cdot\dot{\mathbf{D}} + \mathbf{B}\cdot\mathbf{H} = \sum_{kl} E_k\varepsilon_{kl}\dot{E}_l + \frac{1}{2}\frac{d}{dt}(\mu H^2). \tag{6}$$

Если разделить обе части этого равенства на 4π, то второй член в правой части будет представлять скорость изменения магнитной энергии в единице объема, но первый его член окажется скоростью изменения плотности электрической энергии лишь при условии

$$\frac{1}{4\pi}\sum_{kl} E_k\varepsilon_{kl}\dot{E}_l = \frac{dw_e}{dt} = \frac{1}{8\pi}\sum_{kl}\varepsilon_{kl}(E_k\dot{E}_l + E_l\dot{E}_k), \tag{7}$$

т. е. при

$$\sum_{kl}\varepsilon_{kl}(E_k\dot{E}_l - \dot{E}_k E_l) = 0.$$

Здесь различие индексов k и l фиктивно, так как оба они принимают одни и те же значения (x, y, z). Следовательно, наше выражение не изменится, если переставить k и l во втором члене, что дает

$$\sum_{kl}(\varepsilon_{kl} - \varepsilon_{lk})E_k\dot{E}_l = 0.$$

Поскольку такое условие должно выполняться при любом значении поля, отсюда следует, что

$$\varepsilon_{kl} = \varepsilon_{lk}. \tag{8}$$

Это означает, что *тензор диэлектрической проницаемости должен быть симметричным*. Из девяти его компонент только шесть независимы. Обратно, условие (8) достаточно, чтобы обеспечить справедливость уравнения (7), и мы получаем теорему, выражающую закон сохранения энергии в дифференциальной форме («гидродинамическое уравнение непрерывности» (1.1.43)), т. е.

$$-\operatorname{div}\mathbf{S} = \frac{dw}{dt} \qquad (w = w_e + w_m). \tag{9}$$

Симметричность тензора ε позволяет привести выражение для электрической энергии w_e к такой форме, при которой сохраняются лишь квадраты компонент поля и отсутствуют их произведения. Рассмотрим в пространстве x, y, z поверхность второго порядка

$$\varepsilon_{xx}x^2 + \varepsilon_{yy}y^2 + \varepsilon_{zz}z^2 + 2\varepsilon_{yz}yz + 2\varepsilon_{xz}xz + 2\varepsilon_{xy}xy = \text{const}. \tag{10}$$

Левая часть уравнения (10) должна иметь положительную квадратичную форму, потому что при замене x, y и z на компоненты вектора **E** она становится равной $8\pi w_e$, а энергия w_e должна быть положительной для любого значения вектора поля. Поэтому уравнение (10) описывает эллипсоид, и его всегда можно привести к главным осям эллипсоида; таким образом, существует система координат, связанная с кристаллом, в которой уравнение эллипсоида имеет вид

$$\varepsilon_x x^2 + \varepsilon_y y^2 + \varepsilon_z z^2 = \text{const}. \tag{11}$$

В этой системе *главных диэлектрических осей* материальные уравнения и выражение для электрической энергии принимают простую форму, а именно:

$$D_x = \varepsilon_x E_x, \qquad D_y = \varepsilon_y E_y, \qquad D_z = \varepsilon_z E_z, \tag{12}$$

$$w_e = \frac{1}{8\pi}(\varepsilon_x E_x^2 + \varepsilon_y E_y^2 + \varepsilon_z E_z^2) = \frac{1}{8\pi}\left(\frac{D_x^2}{\varepsilon_x} + \frac{D_y^2}{\varepsilon_y} + \frac{D_z^2}{\varepsilon_z}\right). \tag{13}$$

Величины ε_x, ε_y, ε_z называются *главными диэлектрическими проницаемостями*. Из приведенных выше формул непосредственно следует, что **D** и **E** всегда имеют различные направления, если только направление вектора **E** не совпадает с одной из главных осей или все главные диэлектрические проницаемости не равны друг другу. В последнем случае ($\varepsilon_x = \varepsilon_y = \varepsilon_z$) эллипсоид вырождается в сферу.

Здесь необходимо сделать замечание о влиянии дисперсии. Напомним, что в случае изотропных сред диэлектрическая проницаемость не является постоянной вещества, а зависит от частоты, и точно так же в анизотропной среде шесть компонент тензора диэлектрической проницаемости ε_{kl} изменяются с изменением частоты. Поэтому меняются не только значения главных диэлектрических проницаемостей ε_x, ε_y, ε_z, но и направления главных осей. Это явление известно как *дисперсия осей*. Однако оно может возникать лишь в тех кристаллических структурах, симметрия которых не позволяет выделить предпочтительный ортогональный триплет направлений; т. е. в кристаллах моноклинной и триклинной систем *) (см. п. 14.3.1).

Можно не учитывать дисперсию, если ограничиться рассмотрением монохроматических волн; тогда величины ε_{kl} являются постоянными, зависящими лишь от свойств вещества.

§ 14.2. Структура монохроматической плоской волны в анизотропной среде

14.2.1. Фазовая и лучевая скорости. Для монохроматической плоской волны с угловой частотой $\omega = 2\pi\nu$, которая распространяется со скоростью c/n в направлении единичного вектора нормали **s**, векторы **E**, **D**, **H** и **B** пропорциональны (в комплексной записи) $\exp\left[i\omega\left(\frac{n}{c}(\mathbf{r}\cdot\mathbf{s})-t\right)\right]$. Заметим сразу же, что в дополнение к *фазовой скорости* (или *скорости по нормали*) c/n нам придется ввести еще *лучевую скорость* (или *скорость энергии*), поскольку, как будет показано далее, в анизотропной среде скорость и направление распространения энергии в общем случае отличаются от скорости волны и направления волновой нормали.

Для такого гармонического поля операция $\partial/\partial t$ всегда эквивалентна умножению на $-i\omega$, а операция $\partial/\partial x$ — умножению на $i\omega n s_x/c$. В частности, имеем

$$\dot{\mathbf{E}} = -i\omega\mathbf{E}, \qquad \operatorname{rot}\mathbf{E} = i\omega\frac{n}{c}\mathbf{s}\times\mathbf{E}. \tag{1}$$

В области, не содержащей токов, т. е. там, где

$$\operatorname{rot}\mathbf{H} - \frac{1}{c}\dot{\mathbf{D}} = 0, \qquad \operatorname{rot}\mathbf{E} + \frac{1}{c}\dot{\mathbf{B}} = 0, \tag{2}$$

уравнения Максвелла принимают вид

$$n\mathbf{s}\times\mathbf{H} = -\mathbf{D}, \qquad n\mathbf{s}\times\mathbf{E} = \mu\mathbf{H}; \tag{3}$$

здесь использовано соотношение $\mathbf{B} = \mu\mathbf{H}$. Исключая из уравнений (3) **H** и используя хорошо известное векторное тождество, получим

$$\mathbf{D} = -\frac{n^2}{\mu}\mathbf{s}\times(\mathbf{s}\times\mathbf{E}) = \frac{n^2}{\mu}[\mathbf{E}-\mathbf{s}(\mathbf{s}\cdot\mathbf{E})] = \frac{n^2}{\mu}\mathbf{E}_\perp. \tag{4}$$

Здесь $\mathbf{E}_\perp$ обозначает векторную компоненту **E**, перпендикулярную к **s** и расположенную в плоскости векторов **E** и **s** (рис. 14.1).

Из уравнений (3) видно, что вектор **H** (а следовательно, и **B**) перпендикулярен к векторам **E**, **D** и **s**, которые поэтому должны быть *компланарны*. Кроме того, вектор **D** должен быть ортогонален к **s**. Таким образом, как и в изотроп-

*) Дисперсия осей особенно заметна в инфракрасной области спектра (см. [1, 2]).

ной среде, векторы **H** и **D** перпендикулярны к направлению распространения **s**, а **E** составляет с ним некоторый угол, отличный от прямого. Рис. 14.1 показывает относительное расположение этих векторов, а также единичного вектора, направление которого совпадает с направлением лучевого вектора **S**. Этот единичный вектор перпендикулярен к **E** и **H** и обозначен символом **t**. Угол между **E** и **D**, равный углу между **s** и **t**, обозначим через α. Мы видим, что *векторы* **D**, **H** *и* **s**, *с одной стороны, и векторы* **E**, **H** *и* **t** *— с другой, образуют ортогональные тройки векторов с общим вектором* **H**, повернутые друг относительно друга на угол α. Таким образом, *в кристалле, вообще говоря, энергия распространяется не в направлении нормали к волне.* Вместе с тем теорема равенства плотностей электрической и магнитной энергий по-прежнему сохраняет свою справедливость. Это следует из уравнения (3), так как *)

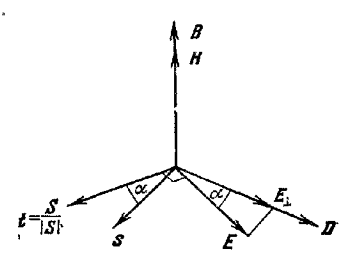


Рис 14 1. Взаимное расположение волновой нормали, векторов поля и вектора потока энергии в электрически анизотропной среде.

$$w_e=\frac{1}{8\pi}\mathbf{E}\cdot\mathbf{D}=-\frac{n}{8\pi}\mathbf{E}\cdot(\mathbf{s}\times\mathbf{H}),$$
$$w_m=\frac{1}{8\pi}\mathbf{B}\cdot\mathbf{H}=\frac{n}{8\pi}(\mathbf{s}\times\mathbf{E})\cdot\mathbf{H}. \qquad (5)$$

Согласно хорошо известным свойствам смешанного произведения правые части обоих уравнений равны между собой. Кроме того, они равны $n(\mathbf{E}\times\mathbf{H})\cdot\mathbf{s}/8\pi$, так что для полной энергии $w=w_e+w_m$ имеем

$$w=\frac{n}{c}\mathbf{S}\cdot\mathbf{s}. \qquad (6)$$

Необходимо установить различие между фазовой скоростью и скоростью распространения энергии. Направление *фазовой скорости* совпадает с направлением единичного вектора **s**, а величина ее равна

$$v_p=\frac{c}{n}. \qquad (7)$$

Направление *лучевой скорости* совпадает с направлением вектора Пойнтинга **S**, т. е. с направлением единичного вектора **t**. Величина ее v_r численно равна отношению энергии, которая протекает в единицу времени через единичную площадку, перпендикулярную к направлению потока, к энергии единицы объема. Согласно теореме (14.1.9) имеем

$$v_r=\frac{S}{w}. \qquad (8)$$

Из последних трех соотношений находим

$$v_p=v_r\mathbf{t}\cdot\mathbf{s}=v_r\cos\alpha, \qquad (9)$$

т. е. *фазовая скорость равна проекции лучевой скорости на направление волновой нормали.*

Следует отметить, что поскольку лучевая скорость определяется через вектор Пойнтинга, для нее также характерна известная доля неопределенности (см. § 1.1). Тем не менее это полезная величина, хотя в отличие от фазовой скорости она не имеет такого явного физического смысла.

Если **E** и **D** известны (например, **E** задано, а **D** определяется из уравнений (14.1.1)), то можно определить показатель преломления n и вектор волновой

*) Поскольку мы теперь имеем дело с квадратичными функциями, в выражения (5) и (6) входят вещественные векторы поля, а не соответствующие комплексные векторы (см. стр. 38).

нормали $\mathbf{s}$. Прежде всего, так как $\mathbf{E}_\perp$ — векторная компонента $\mathbf{E}$ в направлении $\mathbf{D}$, т. е.

$$\mathbf{E}_\perp = \left(\mathbf{E}\cdot\frac{\mathbf{D}}{D}\right)\frac{\mathbf{D}}{D}, \tag{10}$$

то из (4) получим

$$n^2 = \frac{\mu \mathbf{D}}{\mathbf{E}_\perp} = \frac{\mu \mathbf{D}^2}{\mathbf{E}\cdot\mathbf{D}}. \tag{11}$$

Далее, поскольку единичный вектор $\mathbf{s}$ перпендикулярен к $\mathbf{D}$ и компланарен с $\mathbf{D}$ и $\mathbf{E}$, его можно представить в виде

$$\mathbf{s} = \frac{\mathbf{E}-\mathbf{E}_\perp}{|\mathbf{E}-\mathbf{E}_\perp|} = \frac{\mathbf{E}-\dfrac{(\mathbf{E}\cdot\mathbf{D})\,\mathbf{D}}{\mathbf{D}^2}}{\sqrt{\mathbf{E}^2-\dfrac{(\mathbf{E}\cdot\mathbf{D})^2}{\mathbf{D}^2}}} = \frac{\mathbf{D}^2\mathbf{E}-(\mathbf{E}\cdot\mathbf{D})\,\mathbf{D}}{\sqrt{\mathbf{D}^2\,[\mathbf{E}^2\mathbf{D}^2-(\mathbf{E}\cdot\mathbf{D})^2]}}. \tag{12}$$

По аналогии с показателем преломления n можно также определить *лучевой*, или *энергетический, показатель* n_r посредством формулы

$$n_r = \frac{c}{v_r}. \tag{13}$$

Из (7) и (9) найдем

$$n_r = n\cos\alpha. \tag{14}$$

Сейчас мы покажем, что лучевой показатель n_r и единичный вектор $\mathbf{t}$, расположенный в направлении распространения энергии, определяются формулами, аналогичными формулам (11) и (12). Используя (11), (14) и соотношение $\mathbf{E}\cdot\mathbf{D} = ED\cos\alpha$, получим

$$n_r^2 = \frac{\mu\,(\mathbf{E}\cdot\mathbf{D})}{\mathbf{E}^2}. \tag{15}$$

Единичный вектор $\mathbf{t}$ перпендикулярен к $\mathbf{E}$ и компланарен с $\mathbf{E}$ и $\mathbf{D}$ и поэтому должен определяться (с точностью до знака) формулами, которые получаются перестановкой $\mathbf{E}$ и $\mathbf{D}$ в (12). Следовательно,

$$-\mathbf{t} = \frac{\mathbf{E}^2\mathbf{D}-(\mathbf{E}\cdot\mathbf{D})\,\mathbf{E}}{\sqrt{\mathbf{E}^2\,[\mathbf{E}^2\mathbf{D}^2-(\mathbf{E}\cdot\mathbf{D})^2]}}. \tag{16}$$

Отрицательный знак при $\mathbf{t}$ слева обеспечивает соответствие между направлениями векторов $\mathbf{s}$ и $\mathbf{t}$ и векторов $\mathbf{E}$ и $\mathbf{D}$, как показано на рис. 14.1.

Как (12), так и (16) сводятся к неопределенности при совпадении направлений $\mathbf{E}$ и $\mathbf{D}$, т. е. когда вектор $\mathbf{E}$ направлен вдоль одной из главных осей кристалла. Этого следовало ожидать, поскольку в данном случае направления $\mathbf{s}$ и $\mathbf{t}$ неопределенны и известно лишь, что они должны быть перпендикулярны к $\mathbf{E}$.

Можно также выразить величину вектора Пойнтинга через $\mathbf{E}$ и $\mathbf{D}$. Учитывая, что для плоской волны $w = 2w_e$, получим из (8), (13) и (15)

$$S = v_r w = \frac{c}{n_r}\,\frac{\mathbf{E}\cdot\mathbf{D}}{4\pi} = \frac{c}{4\pi\sqrt{\mu}}\,E\sqrt{\mathbf{E}\cdot\mathbf{D}}. \tag{17}$$

Как мы видим, для изотропной среды эта формула согласуется с уравнениями (1.4.8) и (1.4.9).

14.2.2. Формулы Френеля для распространения света в кристаллах. Формулы, полученные в п. 14.2.1, являются следствием одних лишь уравнений Максвелла и поэтому не зависят от свойств среды. Объединим их теперь с материальными уравнениями (14.1.1).

Выберем в качестве осей координат главные диэлектрические оси. Тогда соотношения (14.1.1) примут более простую форму (14.1.12), и, подставляя $\mathbf{D}$ в (4), получим

$$\mu\varepsilon_k E_k = n^2\,[E_k - s_k\,(\mathbf{E}\cdot\mathbf{s})] \qquad (k = x,\ y,\ z). \tag{18}$$

Уравнения (18), которые представляют собой три линейных однородных уравнения для E_x, E_y и E_z, допускают нетривиальное решение только тогда, когда соответствующий определитель обращается в нуль. Это означает, что между показателем преломления n, вектором $\mathbf{s}$ (s_x, s_y, s_z) и главными диэлектрическими проницаемостями ε_x, ε_y и ε_z должно выполняться определенное соотношение, которое можно получить, записав уравнение (18) в виде

$$E_k = \frac{n^2 s_k (\mathbf{E}\cdot\mathbf{s})}{n^2 - \mu\varepsilon_k}, \tag{19}$$

умножив его на s_k и сложив три полученных уравнения. Разделив окончательное выражение на общий множитель $(\mathbf{E}\cdot\mathbf{s})$, найдем

$$\frac{s_x^2}{n^2-\mu\varepsilon_x} + \frac{s_y^2}{n^2-\mu\varepsilon_y} + \frac{s_z^2}{n^2-\mu\varepsilon_z} = \frac{1}{n^2}. \tag{20}$$

Это соотношение можно представить в несколько ином виде. Умножим обе части (20) на n^2 и вычтем $s_x^2+s_y^2+s_z^2=1$. Затем, умножив получившееся выражение на $-n^2$, найдем

$$\frac{s_x^2}{1/n^2-1/\mu\varepsilon_x} + \frac{s_y^2}{1/n^2-1/\mu\varepsilon_y} + \frac{s_z^2}{1/n^2-1/\mu\varepsilon_z} = 0. \tag{21}$$

Определим три *главные скорости распространения* с помощью формул *)

$$v_x = \frac{c}{\sqrt{\mu\varepsilon_x}}, \qquad v_y = \frac{c}{\sqrt{\mu\varepsilon_y}}, \qquad v_z = \frac{c}{\sqrt{\mu\varepsilon_z}}. \tag{22}$$

Если для фазовой скорости использовать выражение (7), то (19) и (21) принимают вид

$$E_k = \frac{v_k^2}{v_k^2 - v_p^2} s_k (\mathbf{E}\cdot\mathbf{s}) \qquad (k = x,\ y,\ z), \tag{23}$$

$$\frac{s_x^2}{v_p^2-v_x^2} + \frac{s_y^2}{v_p^2-v_y^2} + \frac{s_z^2}{v_p^2-v_z^2} = 0. \tag{24}$$

Уравнения (20), (21) и (24) являются эквивалентными формами *уравнения волновых нормалей Френеля*. Это уравнение квадратично относительно v_p^2, что легко показать, умножив (24) на произведение знаменателей. Таким образом, каждому направлению $\mathbf{s}$ соответствуют две фазовые скорости v_p. (Два значения $\pm v_p$, соответствующие любому значению v_p^2, считаются одним, так как отрицательное значение, очевидно, принадлежит противоположному направлению распространения $-\mathbf{s}$.) Для каждого из двух значений v_p из уравнений (23) можно определить отношения $E_x : E_y : E_z$; соответствующие отношения, содержащие вектор $\mathbf{D}$, можно затем получить из (14.1.12). Так как эти отношения вещественны, поля $\mathbf{E}$ и $\mathbf{D}$ *линейно поляризованы*. Таким образом, мы получили важный результат, а именно: *структура анизотропной среды допускает распространение в любом данном направлении двух монохроматических плоских волн, линейно поляризованных в двух разных направлениях и обладающих различными скоростями*. Позднее будет показано, что два направления вектора электрического смещения $\mathbf{D}$, соответствующие данному направлению распространения $\mathbf{s}$, перпендикулярны друг к другу.

Покажем, что аналогичную формулу можно вывести и для лучевой скорости v_r. Это легко сделать, показав сначала, что справедливо соотношение, аналогичное (4), в котором $\mathbf{D}$ и $\mathbf{s}$ заменено на $\mathbf{E}$ и $\mathbf{t}$, и наоборот. Удобно ввести

*) Отметим, что v_x, v_y, v_z не являются компонентами вектора и определяются лишь относительно главных осей.

вектор $\mathbf{D}_\perp$, который определяется как векторная компонента $\mathbf{D}$, перпендикулярная к $\mathbf{t}$ и лежащая в плоскости векторов $\mathbf{D}$ и $\mathbf{t}$. Этот вектор, очевидно, равен

$$\mathbf{D}_\perp = \mathbf{D} - \mathbf{t}(\mathbf{D}\cdot\mathbf{t}). \tag{25}$$

Так как электрический вектор $\mathbf{E}$ тоже перпендикулярен к $\mathbf{t}$ и компланарен с $\mathbf{D}$ и $\mathbf{t}$ (см. рис. 14.1), то $\mathbf{D}_\perp$ параллелен $\mathbf{E}$ и, значит, его можно представить в виде

$$\mathbf{D}_\perp = \left(\mathbf{D}\cdot\frac{\mathbf{E}}{E}\right)\frac{\mathbf{E}}{E} = \frac{n_r^2}{\mu}\mathbf{E}, \tag{26}$$

где использовано выражение (15). Из последних двух соотношений следует, что

$$\mathbf{E} = \frac{\mu}{n_r^2}[\mathbf{D} - \mathbf{t}(\mathbf{D}\cdot\mathbf{t})] = \frac{\mu}{n_r^2}\mathbf{D}_\perp. \tag{27}$$

Это уравнение аналогично уравнению (4) и формально его можно получить при взаимозамене $\mathbf{E}$ и $\mathbf{D}$, n и $1/n_r$, μ и $1/\mu$ и $\mathbf{s}$ и $-\mathbf{t}$. Из основных уравнений вытекает совершенно общее *правило взаимного соответствия*.

Расположим все интересующие нас переменные в два ряда:

$$\left.\begin{array}{llllllllllllll} \mathbf{E}, & \mathbf{D}, & \mathbf{s}, & \mathbf{t}, & c, & \mu, & v_p, & n, & \varepsilon_x, & \varepsilon_y, & \varepsilon_z, & v_x, & v_y, & v_z, \\ \mathbf{D}, & \mathbf{E}, & -\mathbf{t}, & -\mathbf{s}, & \frac{1}{c}, & \frac{1}{\mu}, & \frac{1}{v_r}, & \frac{1}{n_r}, & \frac{1}{\varepsilon_x}, & \frac{1}{\varepsilon_y}, & \frac{1}{\varepsilon_z}, & \frac{1}{v_x}, & \frac{1}{v_y}, & \frac{1}{v_z}. \end{array}\right\} \tag{28}$$

Тогда, *если в любом соотношении, которое связывает величины, приведенные в одном ряду, заменить все параметры соответствующими параметрами из другого ряда, то полученное соотношение также будет справедливо.*

Применив это правило к уравнению волновых нормалей Френеля (24), мы немедленно получим искомое *лучевое уравнение*

$$\frac{t_x^2}{1/v_r^2 - 1/v_x^2} + \frac{t_y^2}{1/v_r^2 - 1/v_y^2} + \frac{t_z^2}{1/v_r^2 - 1/v_z^2} = 0. \tag{29}$$

Конечно, это уравнение можно записать в форме, аналогичной (20) и (21). Уравнение (29), как и (24), квадратично относительно v_r^2, и для каждого направления луча $\mathbf{t}$ (t_x, t_y, t_z) дает две возможные лучевые скорости v_r. Соответствующее направление вектора $\mathbf{D}$ можно определить, решая при каждом значении v_r уравнение, эквивалентное (23), а именно

$$D_k = -\frac{v_r^2}{v_k^2 - v_r^2}t_k(\mathbf{D}\cdot\mathbf{t}) \qquad (k = x,\ y,\ z). \tag{30}$$

Затем, используя (14.1.12), можно найти направления обоих векторов $\mathbf{E}$ (которые, как мы видели, перпендикулярны к $\mathbf{t}$).

Как правило, задается лишь один из векторов $\mathbf{s}$ или $\mathbf{t}$; поэтому желательно вывести соотношения, с помощью которых можно было бы прямо найти неизвестный вектор. Из рис. 14.1 имеем

$$\mathbf{E}\cdot\mathbf{s} = E_\perp \operatorname{tg}\alpha, \qquad \mathbf{D}\cdot\mathbf{t} = -D\sin\alpha. \tag{31}$$

Но, согласно (4), $D = n^2E_\perp/\mu$. Следовательно, используя (7) и (9), находим

$$\mathbf{D}\cdot\mathbf{t} = -\frac{n^2}{\mu}E_\perp\sin\alpha = -\frac{n^2}{\mu}\mathbf{E}\cdot\mathbf{s}\cos\alpha = -\frac{1}{\mu}\frac{c^2}{v_pv_r}\mathbf{E}\cdot\mathbf{s}. \tag{32}$$

Подстановка (32) в (30) дает

$$D_k = \varepsilon_kE_k = \frac{1}{\mu}\frac{c^2v_r}{v_p(v_k^2 - v_r^2)}t_k(\mathbf{E}\cdot\mathbf{s}). \tag{33}$$

Сравнивая (33) с (23) и учитывая, что $\mu\varepsilon_kv_k^2 = c^2$, получим

$$\frac{v_ps_k}{v_k^2 - v_p^2} = \frac{v_rt_k}{v_k^2 - v_r^2}. \tag{34}$$

Решая относительно t_k, найдем

$$t_k = \frac{v_p}{v_r}\frac{v_k^2 - v_r^2}{v_k^2 - v_p^2} s_k, \tag{35}$$

так что

$$v_r t_k - v_p s_k = v_p s_k \frac{v_r^2 - v_p^2}{v_p^2 - v_k^2}. \tag{36}$$

Возводя в квадрат и складывая три уравнения (36), а затем используя соотношение (9), т. е. $\mathbf{s}\cdot\mathbf{t} = v_p/v_r$, получим

$$v_r^2 - v_p^2 = v_p^2 (v_r^2 - v_p^2)^2 \left[\left(\frac{s_x}{v_p^2 - v_x^2}\right)^2 + \left(\frac{s_y}{v_p^2 - v_y^2}\right)^2 + \left(\frac{s_z}{v_p^2 - v_z^2}\right)^2\right]. \tag{37}$$

Следовательно, мы можем написать

$$g^2 \equiv v_p^2 (v_r^2 - v_p^2) = \frac{1}{\left(\frac{s_x}{v_p^2 - v_x^2}\right)^2 + \left(\frac{s_y}{v_p^2 - v_y^2}\right)^2 + \left(\frac{s_z}{v_p^2 - v_z^2}\right)^2}. \tag{38}$$

Это соотношение выражает v_r через $\mathbf{s}$, так как зависимость v_p от $\mathbf{s}$ уже известна из уравнения Френеля (24). Определив таким образом v_r, получим из уравнения (35) единичный вектор $\mathbf{t}$ как функцию $\mathbf{s}$. Используя выражение для g, можно представить уравнение (35) в виде

$$t_k = \frac{s_k}{v_p v_r}\left(v_p^2 + \frac{g^2}{v_p^2 - v_k^2}\right) \qquad (k = x,\ y,\ z). \tag{39}$$

Так как каждому вектору $\mathbf{s}$ в общем случае соответствуют две фазовые скорости v_p, то для каждого направления волновой нормали имеется два направления луча *). Однако в некоторых кристаллах (двухосные кристаллы, см. п. 14.3.1) существуют два особых направления, которым вследствие исчезновения знаменателей в (39) соответствует *бесконечное* число лучей. Существуют также два особых направления луча, каждому из которых соответствует бесконечное число направлений волновых нормалей. Эти специальные случаи обусловливают интересное явление (коническая рефракция), которое будет рассмотрено в п. 14.3.4.

14.2.3. Геометрические построения для определения скоростей распространения и направлений колебаний. Многие результаты, относящиеся к фазовой и лучевой скоростям и к направлениям колебаний, можно проиллюстрировать с помощью некоторых геометрических построений.

а. Эллипсоид волновых нормалей. Согласно уравнениям (14.1.13) компоненты вектора $\mathbf{D}$ при заданной плотности энергии $w = 2w_e$ удовлетворяют соотношению

$$\frac{D_x^2}{\varepsilon_x} + \frac{D_y^2}{\varepsilon_y} + \frac{D_z^2}{\varepsilon_z} = C \qquad (C = 8\pi w = \mathbf{E}\cdot\mathbf{D}). \tag{40}$$

Заменим $D_x/\sqrt{C}$, $D_y/\sqrt{C}$ и $D_z\sqrt{C}$ на x, y и z и будем рассматривать последние как декартовы координаты в пространстве. Тогда

$$\frac{x^2}{\varepsilon_x} + \frac{y^2}{\varepsilon_y} + \frac{z^2}{\varepsilon_z} = 1. \tag{41}$$

Это уравнение описывает эллипсоид, полуоси которого равны квадратному корню из главных диэлектрических проницаемостей и совпадают по направлению с главными диэлектрическими осями. Мы назовем такой эллипсоид *эллипсоидом волновых нормалей*, употребив это название вместо широко используемого, но довольно неопределенного термина «оптическая индикатриса» (он известен также как эллипсоид индексов).

*) Положение лучей, которые соответствуют данной нормали, подробно рассматривается для случая двухосных кристаллов в книге М. Борна [3].

Если воспользоваться эллипсоидом нормалей, то обе фазовые скорости v_p и оба направления колебаний $\mathbf{D}$, соответствующие данному направлению волновой нормали $\mathbf{s}$, можно найти следующим образом. Через начало координат проведем плоскость, перпендикулярную к $\mathbf{s}$. Сечение эллипсоида нормалей такой плоскостью представляет собой эллипс, направление главных осей которого указывает направление колебаний вектора $\mathbf{D}$, а длины полуосей обратно пропорциональны соответствующим фазовым скоростям v_p (рис. 14.2).

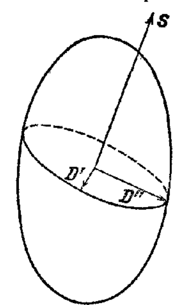


Рис. 14.2. Эллипсоид волновых нормалей.

Построение направлений колебаний векторов D, принадлежащих волновой нормали s.

Для получения этого результата рассмотрим два уравнения, которые определяют наш эллипс:

$$xs_x + ys_y + zs_z = 0, \tag{42}$$

$$\frac{x^2}{\varepsilon_x} + \frac{y^2}{\varepsilon_y} + \frac{z^2}{\varepsilon_z} = 1. \tag{43}$$

Так как по определению главные оси эллипса служат его наименьшим и наибольшим диаметрами, мы можем определить их, находя экстремумы величины

$$r^2 = x^2 + y^2 + z^2 \tag{44}$$

с дополнительными условиями (42) и (43). Это мы сделаем методом неопределенных множителей Лагранжа *). Введем два множителя $2\lambda_1$ и λ_2 и сконструируем функцию

$$F = x^2 + y^2 + z^2 + 2\lambda_1 (xs_x + ys_y + zs_z) + \lambda_2 \left(\frac{x^2}{\varepsilon_x} + \frac{y^2}{\varepsilon_y} + \frac{z^2}{\varepsilon_z}\right). \tag{45}$$

Тогда наша задача сводится к определению экстремума функции F без дополнительных условий. Необходимое условие экстремума функции F состоит в равенстве нулю ее производных по x, y и z, т. е.

$$x + \lambda_1 s_x + \frac{\lambda_2 x}{\varepsilon_x} = 0, \qquad y + \lambda_1 s_y + \frac{\lambda_2 y}{\varepsilon_y} = 0, \qquad z + \lambda_1 s_z + \frac{\lambda_2 z}{\varepsilon_z} = 0. \tag{46}$$

Умножая эти уравнения соответственно на x, y и z и складывая, получим, учитывая (42) и (43),

$$r^2 + \lambda_2 = 0. \tag{47}$$

Теперь, умножая уравнения (46) на s_x, s_y и s_z и складывая, найдем, снова учитывая (42),

$$\lambda_1 + \lambda_2 \left(\frac{xs_x}{\varepsilon_x} + \frac{ys_y}{\varepsilon_y} + \frac{zs_z}{\varepsilon_z}\right) = 0. \tag{48}$$

Подстановка в (46) λ_1 и λ_2 из (47) и (48) дает

$$x\left(1 - \frac{r^2}{\varepsilon_x}\right) + s_x r^2 \left(\frac{xs_x}{\varepsilon_x} + \frac{ys_y}{\varepsilon_y} + \frac{zs_z}{\varepsilon_z}\right) = 0 \tag{49}$$

и еще два аналогичных уравнения. Для заданного $\mathbf{s}$ это три однородных уравнения относительно x, y и z. Они совместны только тогда, когда соответствующий детерминант обращается в нуль, что дает алгебраическое уравнение для r^2. Легко, однако, заметить, что уравнения (49) отличаются от уравнений (18) лишь обозначениями. Если мы заменим x на $D_x/\sqrt{C}$, x/ε_x на $E_x/\sqrt{C}$ и r^2 на $\mathbf{D}^2/C = \mathbf{D}^2/(\mathbf{E}\cdot\mathbf{D}) = n^2/\mu$ (в согласии с (11)), то (49) примет вид

$$\mu D_x = n^2 [E_x - s_x (\mathbf{E}\cdot\mathbf{s})], \tag{50}$$

что вместе с двумя аналогичными уравнениями идентично (18).

*) Полное описание этого метода см., например, в [4].

Таким образом, мы нашли, что корни определительного уравнения для $n = c/v_p$ (оно, как мы видели, квадратично) пропорциональны длинам r полуосей эллиптического сечения, перпендикулярного к $\mathbf{s}$. Кроме того, два возможных направления вектора $\mathbf{D}$ совпадают с направлениями этих осей. Так как оси эллипса взаимно перпендикулярны, то мы получили следующий важный результат: *направления колебаний двух векторов* $\mathbf{D}$, *соответствующих заданному направлению распространения* $\mathbf{s}$, *взаимно перпендикулярны*. Обозначим два направления $\mathbf{D}$, которые соответствуют данному направлению волновой нормали $\mathbf{s}$, через $\mathbf{D}'$ и $\mathbf{D}''$; таким образом, $\mathbf{s}$, $\mathbf{D}'$ и $\mathbf{D}''$ образуют ортогональную тройку векторов.

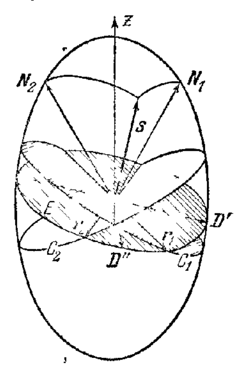


Рис. 14.3 Построение для определения плоскостей колебании (s, D') и (s, D'').

В специальном случае совпадения направления распространения с одной из главных осей эллипсоида нормалей, например, с осью x, экстремумы r равны, согласно нашему построению, длинам двух других полуосей, т. е. $\sqrt{\varepsilon_y}$ и $\sqrt{\varepsilon_z}$. Но мы показали, что экстремумы r равны также $n/\sqrt{\mu} = c/v_p\sqrt{\mu}$. Следовательно, *фазовые скорости волн, которые распространяются в направлении оси* x, *равны* $c/\sqrt{\mu\varepsilon_y}$ *и* $c/\sqrt{\mu\varepsilon_z}$, *т. е. главным скоростям распространения* v_y *и* v_z, *введенным формально с помощью соотношений* (22). Конечно, соответствующие результаты имеют место и для распространения в направлении двух других осей.

Существует и иной способ построения, с помощью которого можно определить направления колебаний. Известно, что у эллипсоида существуют два круговых сечения C_1 и C_2, проходящих через центр, и что нормали к ним $\mathbf{N}_1$ и $\mathbf{N}_2$ компланарны с наибольшей и наименьшей главными осями (z и x) эллипсоида. Направления $\mathbf{N}_1$ и $\mathbf{N}_2$ называются *оптическими осями* *) и будут рассмотрены подробно в п. 14.3.3. Так как C_1 и C_2 — круговые сечения (и имеют одинаковые радиусы), то в направлениях $\mathbf{N}_1$, $\mathbf{N}_2$ существует единственная скорость распространения; при этом $\mathbf{D}$ может иметь *любое* направление, перпендикулярное к $\mathbf{s}$. Пусть E — эллиптическое центральное сечение, перпендикулярное к произвольному единичному вектору нормали $\mathbf{s}$. Плоскость этого сечения пересекает круги C_1 и C_2 вдоль двух радиальных векторов $\mathbf{r}_1$, $\mathbf{r}_2$, которые равны по величине и поэтому должны образовывать равные углы с главными осями сечения E (см. рис. 14.3 и 14.4). Таким образом, искомые направления колебаний являются биссектрисами углов между $\mathbf{r}_1$ и $\mathbf{r}_2$. Но $\mathbf{r}_1$ перпендикулярен к $\mathbf{N}_1$ и $\mathbf{s}$ и поэтому перпендикулярен к плоскости, содержащей $\mathbf{N}_1$ и $\mathbf{s}$; аналогично $\mathbf{r}_2$ перпендикулярен к плоскости, содержащей $\mathbf{N}_2$ и $\mathbf{s}$. Если эти плоскости пересекают плоскость эллипса E вдоль векторов $\mathbf{r}'_1$, $\mathbf{r}'_2$, то главные оси эллипса также должны служить биссектрисами углов между $\mathbf{r}'_1$ и $\mathbf{r}'_2$. Следовательно, *плоскости колебаний вектора электрического смещения*, т. е. плоскости, содержащие $\mathbf{s}$ и $\mathbf{D}'$ или $\mathbf{D}''$, делят пополам внутренний или внешний угол между плоскостями ($\mathbf{N}_1$, $\mathbf{s}$) и ($\mathbf{N}_2$, $\mathbf{s}$) **). Это построение становится неопределенным, если направление $\mathbf{s}$ совпадает с направлением $\mathbf{N}_1$ или $\mathbf{N}_2$.

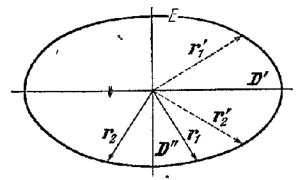


Рис. 14.4. Плоскость E, показанная на рис. 14.3.

*) Точнее, *оптическими осями волновых нормалей*. Соответствующие сечения лучевого эллипсоида (см. ниже) определяют *оптические лучевые оси*.

**) В настоящей главе символ ($\mathbf{a}$, $\mathbf{b}$) означает плоскость, содержащую векторы $\mathbf{a}$ и $\mathbf{b}$.

б. Лучевой эллипсоид. Лучи можно рассматривать таким же образом, как и волновые нормали, если в соответствии с правилом (28) исходить из *лучевого эллипсоида*

$$\varepsilon_x x^2+\varepsilon_y y^2+\varepsilon_z z^2=1. \qquad (51)$$

В частности, центральное сечение этого эллипсоида, перпендикулярное к направлению луча $\mathbf{t}$, является эллипсом, направления главных осей которого указывают два допустимых направления электрического вектора ($\mathbf{E}'$ и $\mathbf{E}''$), а длины полуосей пропорциональны двум соответствующим лучевым скоростям v_r. Таким образом, $\mathbf{t}$, $\mathbf{E}'$ и $\mathbf{E}''$ образуют ортогональную тройку векторов.

в. Поверхность нормалей и лучевая поверхность. Представим себе, что из некоторой точки O внутри кристалла, как из начала координат, в направлении $\mathbf{s}$ откладываются два вектора, длины которых пропорциональны двум соответствующим значениям фазовой скорости. Поскольку вектор $\mathbf{s}$ принимает все возможные направления, концы наших векторов опишут поверхность, состоящую из двух оболочек, называемую *поверхностью волновых нормалей* или, короче, *поверхностью нормалей.*

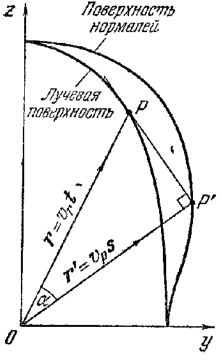


Рис. 14.5. Соотношение между поверхностью нормалей и лучевой поверхностью

Аналогично концы векторов, отложенных из фиксированного начала координат во всех направлениях $\mathbf{t}$ и имеющих длины, пропорциональные соответствующим лучевым скоростям, опишут двухоболочечную поверхность, называемую *лучевой поверхностью*

Эти две поверхности сложнее рассмотренных нами выше поверхностей эллипсоидов Лучевая поверхность — это поверхность четвертого порядка, поверхность нормалей — поверхность шестого порядка *), в чем можно убедиться, обратившись к формулам (24) и (29). Между этими двумя поверхностями существует важное соотношение, которое мы сейчас и получим

Мы показали, что если $\mathbf{E}$ и $\mathbf{D}$ известны, то можно определить как направления $\mathbf{s}$ и $\mathbf{t}$, так и соответствующие скорости v_p и v_r, а следовательно, и соответствующие точки (P и P' на рис 14.5) на обеих описанных выше поверхностях. Пусть $\mathbf{r}$ и $\mathbf{r}'$ — векторы, представляющие эти точки, т. е.

$$\mathbf{r}=v_r\mathbf{t}, \qquad \mathbf{r}'=v_p\mathbf{s}. \qquad (52)$$

Покажем, что приращение вектора $\mathbf{r}$ при небольшом изменении $\mathbf{E}$ или $\mathbf{D}$ перпендикулярно к $\mathbf{r}'$.

Начнем с уравнения (27)

$$\frac{1}{\mu}\mathbf{E}=\frac{1}{n_r^2}[\mathbf{D}-\mathbf{t}(\mathbf{D}\cdot\mathbf{t})]. \qquad (53)$$

Подставляя $\mathbf{t}$ из первого уравнения (52) и полагая $n_r=c/v_r$, получим

$$\frac{c^2}{\mu}\mathbf{E}=r^2\mathbf{D}-\mathbf{r}(\mathbf{D}\cdot\mathbf{r}). \qquad (54)$$

Предположим, что $\mathbf{E}$ изменяется на небольшую величину $\delta\mathbf{E}$. Если $\delta\mathbf{D}$ и $\delta\mathbf{r}$ — соответствующие изменения $\mathbf{D}$ и $\mathbf{r}$, то в соответствии с (54) найдем

$$\frac{c^2}{\mu}\delta\mathbf{E}=2(\mathbf{r}\cdot\delta\mathbf{r})\mathbf{D}+r^2\delta\mathbf{D}-\delta\mathbf{r}(\mathbf{D}\cdot\mathbf{r})-\mathbf{r}(\delta\mathbf{r}\cdot\mathbf{D})-\mathbf{r}(\mathbf{r}\cdot\delta\mathbf{D}). \qquad (55)$$

*) Нельзя ожидать, чтобы поверхность нормалей и лучевая поверхность описывались уравнениями одинаковой степени, так как они не являются взаимно соответствующими друг другу Для построения поверхности, соответствующей поверхности нормалей, необходимо, согласно правилу (28), откладывать векторы длиной $1/v_r$ (а не v_r).

Если обе части этого равенства скалярно умножить на $\mathbf{D}$ и воспользоваться соотношением

$$\mathbf{D}\cdot\delta\mathbf{E}=\varepsilon_x E_x \delta E_x+\varepsilon_y E_y \delta E_y+\varepsilon_z E_z \delta E_z=\mathbf{E}\cdot\delta\mathbf{D}, \tag{56}$$

то мы получим

$$\frac{c^2}{\mu}\mathbf{E}\cdot\delta\mathbf{D}=\delta\mathbf{D}\cdot[r^2\mathbf{D}-\mathbf{r}(\mathbf{D}\cdot\mathbf{r})]+2\delta\mathbf{r}\cdot[\mathbf{r}D^2-\mathbf{D}(\mathbf{D}\cdot\mathbf{r})]. \tag{57}$$

Члены, содержащие множитель $\delta\mathbf{D}$, сокращаются согласно (54), а член с $\delta\mathbf{r}$ можно переписать в виде $2\delta\mathbf{r}\cdot[(\mathbf{D}\times\mathbf{r})\times\mathbf{D}]$. Следовательно, учитывая, что $\mathbf{r}=v_r\,\mathbf{t}$, имеем

$$\delta\mathbf{r}\cdot[(\mathbf{D}\times\mathbf{t})\times\mathbf{D}]=0. \tag{58}$$

Вектор $\mathbf{D}\times\mathbf{t}$ перпендикулярен и к $\mathbf{D}$, и к $\mathbf{t}$, поэтому $(\mathbf{D}\times\mathbf{t})\times\mathbf{D}$ лежит в плоскости векторов $\mathbf{D}$ и $\mathbf{t}$ и перпендикулярен к $\mathbf{D}$. Таким образом, вектор $(\mathbf{D}\times\mathbf{t})\times\mathbf{D}$ параллелен $\mathbf{s}$ (см. § 14.1) и, следовательно,

$$\mathbf{s}\cdot\delta\mathbf{r}=0, \tag{59}$$

т. е. $\delta\mathbf{r}$ перпендикулярен к $\mathbf{s}$, что и доказывает наше утверждение. Отсюда следует, что *плоскость, касательная к лучевой поверхности, всегда перпендикулярна соответствующей волновой нормали.* Рис. 14.5 иллюстрирует это соотношение на плоском сечении. Так как кратчайшее расстояние от начала координат до этой плоскости равно, согласно (9), $v_r\mathbf{t}\cdot\mathbf{s}=v_r\cos\alpha=v_p$, то, следовательно, *поверхность нормалей представляет собой геометрическое место оснований перпендикуляров, опущенных из начала координат на плоскости, касательные к лучевой поверхности, и, обратно, лучевая поверхность является огибающей плоскостей, проведенных через точки поверхности нормалей перпендикулярно радиусам-векторам этих точек.* Если нам известна форма одной из этих поверхностей, указанное соотношение позволит определить форму другой.

Полученный результат можно интерпретировать с физической точки зрения. Рассмотрим не одну волну, а группу плоских волн одинаковой частоты, имеющих слегка различные направления распространения. Волновые нормали $\mathbf{s}$ составляющих волн заполняют телесный угол вокруг «средней волновой нормали» $\mathbf{s}_0$. Предположим, что заметную величину имеют амплитуды лишь тех волн, нормали которых близки к $\mathbf{s}_0$. Допустим, что в момент времени t 0 фазы всех волн в точке O одинаковы; тогда возмущение в ней максимально. Исследуем теперь распространение этого максимума.

Рассмотрим все волновые фронты, которые проходят через точку O в момент $t=0$. Через единицу времени волновой фронт W, распространяющийся со скоростью v_p в направлении $\mathbf{s}$, достигнет такого положения W', что основание перпендикуляра, опущенного на него из точки O, совпадет с концом вектора $v_p\mathbf{s}$. Таким образом, W' — это плоскость, перпендикулярная к соответствующему радиусу-вектору поверхности нормалей. Амплитуда группы волн будет наибольшей в той области, где волны усиливают друг друга, т. е. там, где эта плоскость пересекает плоскости с близкими волновыми нормалями Но такая область должна находиться как раз вблизи огибающей этих плоскостей, т. е. около соответствующей точки $v_r\mathbf{t}$ на лучевой поверхности. Приведенные выше соображения подтверждают, что энергия, переносимая группой, распространяется со скоростью v_r в направлении единичного вектора $\mathbf{t}$.

§ 14.3. Оптические свойства одноосных и двухосных кристаллов

14.3.1. Оптическая классификация кристаллов. Прозрачные кристаллы делятся по своим оптическим свойствам на три различные группы.

Группа I. *Кристаллы, в которых можно выбрать три кристаллографически эквивалентных взаимно ортогональных направления.* Это кристаллы так называемой *кубической* системы. Очевидно, что эквивалентные направления

совпадают с главными диэлектрическими осями, поэтому $\varepsilon_x = \varepsilon_y = \varepsilon_z = \varepsilon$. Тогда $\mathbf{D} = \varepsilon\mathbf{E}$, а кристалл *оптически изотропен* и эквивалентен аморфному телу.

Группа II. *Кристаллы, не принадлежащие к группе I, в которых можно выбрать два или более кристаллографически эквивалентных направления, лежащих в одной плоскости.* Это кристаллы тригональной, тетрагональной и гексагональной систем, причем плоскость, в которой лежат эквивалентные направления, перпендикулярна к осям симметрии третьего, четвертого или шестого порядков. Одна из главных диэлектрических осей должна совпадать с этим выделенным направлением, тогда как для двух других направлений можно выбрать любую взаимно ортогональную пару перпендикулярных к нему прямых. Если выделенное направление принять за ось z, то $\varepsilon_x = \varepsilon_y \neq \varepsilon_z$. Такие кристаллы называют оптически *одноосными*.

Группа III. *Кристаллы, в которых невозможно выбрать два кристаллографически эквивалентных направления.* Такие кристаллы принадлежат к так называемым ромбической, моноклинной и триклинной системам. Здесь $\varepsilon_x \neq \varepsilon_y \neq \varepsilon_z$, а направления диэлектрических осей могут определяться (но могут и не определяться) симметрией (см табл. 14.1) и поэтому могут зависеть от длины волны. Кристаллы этой группы называют оптически *двухосными*.

Таблица 14.1

Кристаллическая система	Диэлектрические оси	Эллипсоид волновых нормалей	Оптическая классификация кристалла
Триклинная	*CCC*	Произвольный эллипсоид	Двухосный
Моноклинная	*CCF*	То же	»
Ромбическая	*FFF*	» »	»
Тригональная Тетрагональная Гексагональная	*FRR*	Сфероид	Одноосный
Кубическая	*RRR*	Сфера	Изотропный

В том, что все кристаллы делятся по своим оптическим свойствам на эти три типа, легко убедиться, рассматривая один из соответствующих эллипсоидов, например эллипсоид волновых нормалей. Очевидно, такой эллипсоид не должен изменяться при операциях симметрии, не меняющих структуру кри-

сталла *). Возможны лишь три случая: эллипсоид может иметь либо (а) все оси разной длины, либо (б) две равные оси и одну не равную им (сфероид, т. е. эллипсоид вращения), либо (в) все оси равной длины (сфера). Эти три возможности соответствуют трем группам (в порядке III, II и I), которые мы только что рассмотрели. Термины «одноосный» и «двухосный» относятся к числу оптических осей эллипсоида, т. е. к числу диаметров эллипсоида, перпендикулярных к его круговому центральному сечению. В общем случае эллипсоид имеет два таких диаметра (двухосные кристаллы), сфероид — один (одноосные кристаллы), а сфера — бесконечное число (изотропные кристаллы).

В табл. 14.1 приведены все возможные случаи. Главные диэлектрические оси, положение которых может зависеть от длины волны (C), показаны двумя тонкими линиями под небольшим углом друг к другу (что указывает на их положение для двух длин волн), фиксированные оси (F) изображены жирными линиями, а оси с произвольным направлением (R) показаны в виде пунктирных линий, заканчивающихся на круге или сфере.

14.3.2. Распространение света в одноосных кристаллах. Начнем с уравнения волновых нормалей Френеля (14.2.24) и запишем его в виде

$$s_x^2(v_p^2-v_y^2)(v_p^2-v_z^2)+s_y^2(v_p^2-v_z^2)(v_p^2-v_x^2)+s_z^2(v_p^2-v_x^2)(v_p^2-v_y^2)=0. \qquad (1)$$

Для оптически одноосных кристаллов с оптической осью вдоль направления z имеем $v_x=v_y$. Обозначив через v_o эту общую скорость и через v_e скорость v_z, получим из (1) **)

$$(v_p^2-v_o^2)\left[(s_x^2+s_y^2)(v_p^2-v_e^2)+s_z^2(v_p^2-v_o^2)\right]=0. \qquad (2)$$

Пусть ϑ — угол, образуемый нормалью $\mathbf{s}$ с осью z; тогда

$$s_x^2+s_y^2=\sin^2\vartheta, \qquad s_z^2=\cos^2\vartheta,$$

и (2) переходит в

$$(v_p^2-v_o^2)\left[(v_p^2-v_e^2)\sin^2\vartheta+(v_p^2-v_o^2)\cos^2\vartheta\right]=0. \qquad (3)$$

Двумя корнями этого уравнения (скажем, v_p' и v_p'') служат

$$(v_p')^2=v_o^2, \qquad (v_p'')^2=v_o^2\cos^2\vartheta+v_e^2\sin^2\vartheta. \qquad (4)$$

Уравнения (4) показывают, что *двумя оболочками поверхности нормалей служат сфера радиуса $v_p'=v_o$ и овалоид — поверхность вращения четвертого порядка*. Таким образом, одной из двух волн, соответствующих любому данному направлению волновой нормали, является *обыкновенная волна*, скорость которой не зависит от направления распространения. Другая — *необыкновенная волна*, скорость которой зависит от угла между направлением волновой нормали и оптической осью. Обе скорости равны лишь при $\vartheta=0$, т. е. когда волновая нормаль направлена вдоль оптической оси.

Когда $v_o>v_e$ (рис. 14.6, *а*), обыкновенная волна распространяется быстрее, чем необыкновенная (исключая направление $\vartheta=0$, когда их скорости равны). Такой кристалл называют *положительным* одноосным кристаллом (например, кварц). Если $v_o<v_e$ (рис. 14.6, *б*), обыкновенная волна распространяется медленнее, чем необыкновенная, и мы называем такой кристалл *отрицательным* одноосным кристаллом (например, исландский шпат).

---

*) Например, кристаллы моноклинной системы характеризуются либо осью второго порядка, параллельной одной из кристаллографических осей, либо зеркальной плоскостью, перпендикулярной ей, либо и тем, и другим. Ясно, что во всех случаях одна из осей этого эллипсоида должна быть параллельна такой кристаллографической оси, которая обеспечивает его сохранение при операциях, не меняющих симметрию кристаллической структуры. Подробное описание кристаллических классов и операций симметрии приводится в книге [5].

**) Здесь индексы o и e означают «обыкновенный» и «необыкновенный», что станет понятным в дальнейшем.

Направления колебаний нетрудно найти обычным способом с помощью эллипсоида волновых нормалей, у которого в данном случае две равные главные оси. Плоскость, в которой лежит волновая нормаль s и оптическая ось Oz, называется *главной плоскостью* (на рис. 14.7 она заштрихована). Эллипсоид

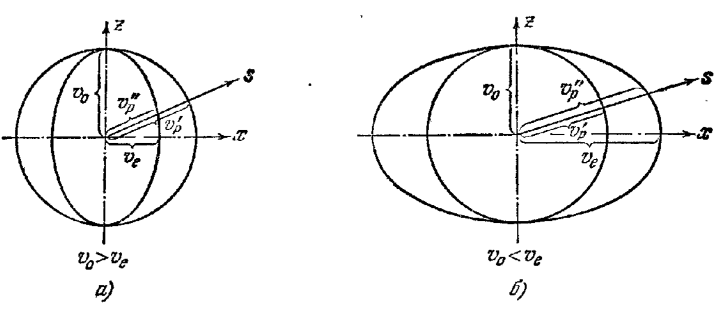


Рис. 14.6. Поверхности нормалей положительного (*а*) и отрицательного (*б*) одноосного кристалла.

симметричен относительно этой плоскости. Отсюда вытекает, что эллиптическое сечение плоскостью, проходящей через O и перпендикулярной к s, симметрично относительно главной плоскости и, следовательно, одна из главных осей эллипса перпендикулярна, а другая параллельна главной плоскости (см. рис. 14.7). Длина полуоси, перпендикулярной к главной плоскости, равна радиусу экваториального круга сфероида, т. е. обратно пропорциональна скорости v_o обыкновенной волны. Мы видим, что *вектор* **D** *обыкновенной волны* (**D**′ на рис. 14.7) *колеблется перпендикулярно к главной плоскости, а вектор необыкновенной волны* (**D**″) *— в главной плоскости.*

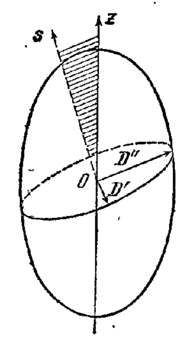


Рис. 14.7. Направления колебаний в одноосном кристалле.

Оптические явления в одноосных кристаллах сыграли значительную роль в истории оптики в связи с вопросом о том, перпендикулярно ли колебание «светового вектора» к плоскости поляризации или параллельно ей. Плоскость поляризации определялась как плоскость падения света, падающего под таким углом, что любая падающая волна превращается при отражении от плоской границы воздух — диэлектрик в линейно поляризованную, т. е. на языке электромагнитной теории как плоскость (**H**, s) (см. стр. 47 и 59). Сегодня не имеет смысла подробно обсуждать этот вопрос *), так как мы знаем, что нет одного-единственного физического понятия, которое можно было бы считать «световым вектором».

14.3.3. Распространение света в двухосных кристаллах. Теперь исследуем главные следствия основного уравнения (1) для общего случая двухосного кристалла. Оно поможет нам наглядно представить поверхность нормалей, если вначале мы рассмотрим сечения этой поверхности тремя координатными плоскостями $x=0$, $y=0$ и $z=0$ нашей исходной системы (главные диэлектрические оси). Для определенности будем считать, что

$$\varepsilon_x < \varepsilon_y < \varepsilon_z \qquad (v_x > v_y > v_z). \tag{5}$$

*) Описание истории вопроса см. в книге [6].

Если в уравнении (1) положить $s_x = 0$, то оно распадется на два, а именно

$$(v_p')^2 = v_x^2, \qquad (v_p'')^2 = v_z^2 s_y^2 + v_y^2 s_z^2. \tag{6}$$

Положим $v_p s_y = y$, $v_p s_z = z$, тогда $v_p^2 = y^2 + z^2$, и уравнения (6) примут вид

$$y^2 + z^2 = v_x^2, \qquad (y^2 + z^2)^2 = v_z^2 y^2 + v_y^2 z^2. \tag{6а}$$

Таким образом, в сечении поверхности нормалей координатной плоскостью $x = 0$ получаются окружность и овал. В сечении каждой из оставшихся двух координатных плоскостей также получаются окружность и овал, причем единственное различие заключается в относительном расположении этих двух кривых. При выборе осей, заданном (5), вся окружность расположена вне овала в плоскости yz и внутри него в плоскости xy. В плоскости zx окружность и овал пересекаются в четырех точках (рис. 14.8). На рис. 14.9 в перспективе показана часть поверхности нормалей, стягиваемая кривыми $A'B'C'N$ и $A''B''C''N$. В общем случае две поверхности пересекаются по кривой, но в нашем случае они имеют лишь четыре общие точки: точку N и соответствующие точки в других квадрантах. Две линии, соединяющие начало координат с каждой из этих точек, являются двумя оптическими осями волновых нормалей. Из геометрической теоремы о числе центральных круговых сечений эллипсоида, упомянутой на стр. 623, следует, что не существует других таких точек и, значит, других оптических осей волновых нормалей Мы подтвердим это прямым расчетом и одновременно выведем неравенство, которое понадобится нам в дальнейшем.

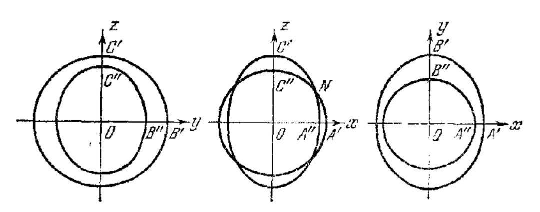


Рис. 14.8. Сечения поверхности нормалей двухосного кристалла.

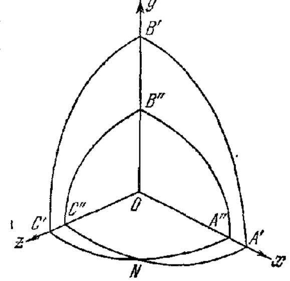


Рис. 14.9. Поверхность нормалей двухосного кристалла.

Пусть

$$v_x^2 = v_y^2 + q_x, \qquad v_z^2 = v_y^2 - q_z, \qquad v_p^2 = v_y^2 + q, \tag{7}$$

где q_x и q_z положительны (см. (5)). При такой замене уравнение (1) примет вид

$$s_x^2 q (q + q_z) + s_y^2 (q + q_z)(q - q_x) + s_z^2 (q - q_x) q = 0 \tag{8а}$$

или

$$q^2 + [s_x^2 q_z + s_y^2 (q_z - q_x) - s_z^2 q_x] \, q - s_y^2 q_x q_z = 0. \tag{8б}$$

Так как постоянный член $-s_y^2 q_x q_z$ не может быть положительным, корни этого уравнения должны быть вещественными. Если мы обозначим их через q' и q'', то найдем

$$q'q'' = -s_y^2 q_x q_z \leqslant 0.$$

Следовательно, q' и q'' должны иметь разные знаки. Пусть $q' \geqslant 0$, а $q'' \leqslant 0$.

Если $q > q_x$ или $q < -q_z$, то все члены слева в уравнении (8а) положительны. Следовательно, значения q должны лежать в интервале $-q_z \leqslant q \leqslant q_x$ и, таким образом,

$$-q_z \leqslant q'' \leqslant 0 \leqslant q' \leqslant q_x. \tag{9}$$

Два корня q' и q'' могут быть равными, только если оба они равны нулю. Тогда,

согласно (8б), мы должны иметь одновременно

$$s_y^2=0, \qquad s_{xz}^2=s_z^2 q_x. \tag{10}$$

Таким образом, мы получили два направления s, для каждого из которых соответствующие две скорости равны, что подтверждает существование двух оптических осей волновых нормалей. Как уже было отмечено ранее, эти оси лежат в плоскости xz. Если β — угол, который образует одна из осей с направлением z, то $s_x=\sin\beta$ и $s_z=\cos\beta$, где, согласно (10),

$$\operatorname{tg}\beta=\frac{s_x}{s_z}=\pm\sqrt{\frac{q_x}{q_z}}=\pm\sqrt{\frac{v_x^2-v_y^2}{v_y^2-v_z^2}}. \tag{11}$$

Таким образом, оптические оси расположены симметрично относительно оси z.

Неравенства (9) можно записать через скорости, т. е.

$$v_z^2\leqslant(v_p'')^2\leqslant v_y^2\leqslant(v_p')^2\leqslant v_x^2. \tag{12}$$

Это соотношение понадобится нам в дальнейшем. Мы видим, что обе фазовые скорости вещественны для любого направления волновой нормали. Такое же заключение вытекает, разумеется, из геометрического построения для фазовых скоростей с привлечением эллипсоида волновых нормалей.

Выражение для двух фазовых скоростей, соответствующих заданному направлению волновой нормали s, принимает очень простую форму, если s определить через углы ϑ_1 и ϑ_2, которые эта нормаль образует с двумя оптическими осями волновых нормалей. Поскольку направляющие косинусы оптических осей равны $\pm\sin\beta$, 0, $\cos\beta$, углы ϑ_1 и ϑ_2 определяются соотношениями

$$\cos\vartheta_1=s_x\sin\beta+s_z\cos\beta, \qquad \cos\vartheta_2=-s_x\sin\beta+s_z\cos\beta. \tag{13}$$

Корни уравнения (8б), выраженные через s_x и s_z, имеют вид

$$q=\pm\frac{1}{2}P\pm\frac{1}{2}\sqrt{\Delta}, \tag{14}$$

где, если воспользоваться тождеством $s_x^2+s_y^2+s_z^2=1$,

$$P=s_x^2q_x-s_z^2q_z+q_z-q_x, \tag{15}$$

$$\Delta=P^2+4s_y^2q_xq_z=(q_x+q_z)^2-2(q_x+q_z)(s_x^2q_x+s_z^2q_z)+(s_z^2q_z-s_x^2q_x)^2. \tag{16}$$

Но из (13) и (11) имеем

$$\cos\vartheta_1\cos\vartheta_2=\frac{q_zs_z^2-q_xs_x^2}{q_x+q_z}, \qquad \cos^2\vartheta_1+\cos^2\vartheta_2=2\frac{q_xs_x^2+q_zs_z^2}{q_x+q_z}. \tag{17}$$

Следовательно, выразив (15) и (16) через ϑ_1 и ϑ_2, получим

$$P=q_z-q_x-(q_z+q_x)\cos\vartheta_1\cos\vartheta_2, \tag{15a}$$

$$\Delta=[(q_z+q_x)\sin\vartheta_1\sin\vartheta_2]^2. \tag{16a}$$

Подставляя (15а) и (16а) в (14) и заменяя q, q_x и q_z их значениями из выражений (7), окончательно находим

$$v_p^2=\frac{1}{2}\left[v_x^2+v_z^2+(v_x^2-v_z^2)\cos(\vartheta_1\pm\vartheta_2)\right]. \tag{18}$$

Хотя в последнее уравнение и не входит v_y, оно неявно содержится в ϑ_1 и ϑ_2, так как эти углы зависят от угла β, являющегося функцией всех трех главных скоростей.

В частном случае одноосного кристалла ϑ_1 и ϑ_2 равны. Легко показать, что тогда уравнение (18) точно переходит в (4).

Подобный анализ применим и к лучам. Начав с лучевого уравнения (14.2.29), мы найдем, что в сечении лучевой поверхности координатной плоскостью $x=0$ получаются две кривые

$$(v_r')^2=v_x^2, \qquad \frac{1}{(v_r'')^2}=\frac{t_y^2}{v_z^2}+\frac{t_z^2}{v_y^2}. \tag{19}$$

Если положить $v_r t_y = y$, $v_r t_z = z$, то $v_r^2 = y^2 + z^2$, и (19) примет вид

$$y^2 + z^2 = v_x^2, \qquad \frac{y^2}{v_z^2} + \frac{z^2}{v_y^2} = 1, \tag{19a}$$

т. е. сечение лучевой поверхности плоскостью $x = 0$ дает окружность и эллипс (а не овал — кривую более общего типа, как в случае (6а)). Каждое из сечений двумя другими координатными плоскостями также представляет собой окружность и эллипс. В соответствии с неравенством (5) в плоскости yz эллипс находится внутри окружности, тогда как в плоскости xy он охватывает окружность. В плоскости xz окружность и эллипс пересекаются в четырех точках. Эти точки указывают положение *лучевых оптических осей* R_1, R_2; их направления определяются соотношениями, соответствующими соотношениям (10), а именно

$$t_y^2 = 0, \qquad t_x^2\left(\frac{1}{v_y^2} - \frac{1}{v_z^2}\right) = t_z^2\left(\frac{1}{v_x^2} - \frac{1}{v_y^2}\right). \tag{20}$$

Угол γ между каждой из этих осей и осью x поэтому определяется выражением

$$\operatorname{tg}\gamma = \frac{t_x}{t_y} = \pm\frac{v_z}{v_x}\sqrt{\frac{v_x^2 - v_y^2}{v_y^2 - v_z^2}} = \pm\frac{v_z}{v_x}\operatorname{tg}\beta. \tag{21}$$

Если $v_z < v_x$, лучевая оптическая ось образует меньший угол с осью z, чем ось нормалей.

14.3.4. Преломление в кристаллах. а. *Двойное лучепреломление.* Рассмотрим плоскую волну, падающую из вакуума на плоскую поверхность Σ анизотропной среды. Эта волна создает прошедшее и отраженное поля. Мы кратко рассмотрим характер прошедшего поля, используя по существу те же рассуждения, что и в случае изотропных тел (см. п. 1.5.1). Ограничимся определением направления распространения возмущения внутри кристалла и не будем исследовать выражений для отношений амплитуд, соответствующих формулам Френеля *).

Пусть $\mathbf{s}$ и $\mathbf{s}'$ — единичные векторы волновой нормали падающей и прошедшей волн соответственно. Вскоре мы увидим, что в общем случае имеются *две* проходящие волны, так что имеются два возможных значения $\mathbf{s}'$. Векторные поля падающей и прошедшей волн являются функциями величин $(t - \mathbf{r}\cdot\mathbf{s}/c)$ и $(t - \mathbf{r}\cdot\mathbf{s}'/v')$ соответственно. Из условия непрерывности поля на границе раздела вытекает, что для любой точки $\mathbf{r}$ на плоскости Σ и для всех моментов времени t справедливо соотношение

$$t - \frac{\mathbf{r}\cdot\mathbf{s}}{c} = t - \frac{\mathbf{r}\cdot\mathbf{s}'}{v'},$$

т. е.

$$\mathbf{r}\cdot\left(\frac{\mathbf{s}'}{v'} - \frac{\mathbf{s}}{c}\right) = 0. \tag{22}$$

Следовательно, вектор $\mathbf{s}'/v' - \mathbf{s}/c$ должен быть перпендикулярен к границе раздела. Допустимые направления волновых нормалей $\mathbf{s}'$ можно определить следующим образом. Из произвольной точки O на плоскости Σ, как из начала координат, во всех направлениях $\mathbf{s}'$ отложим векторы длиной $1/v'$, где v' — фазовая скорость, соответствующая каждому направлению $\mathbf{s}'$ согласно уравнению Френеля (14.2.24). Концы векторов образуют двухоболочечную поверхность, которая отличается от поверхности нормалей тем, что длина каждого радиуса-вектора составляет $1/v'$ вместо v'. Эта поверхность называется *обратной поверхностью волновых нормалей.* Она соответствует лучевой поверхности и поэтому, как и лучевая поверхность, представляет собой поверхность четвертого порядка. Поскольку искомый вектор $\mathbf{s}'/v'$ должен быть таким, чтобы

*) Они обсуждаются, например, в [7].

вектор $\mathbf{s}'/v' - \mathbf{s}/c$ был перпендикулярен к Σ, его конец Q' должен лежать на нормали к Σ, проведенной через конец P вектора $\mathbf{s}/c$. В общем случае нормаль к Σ пересекает обратную поверхность в четырех точках, две из которых расположены с той же стороны поверхности раздела, что и кристалл. Следовательно, мы нашли две нужные точки (Q' и Q'' на рис. 14.10), т. е. два возможных направления волновой нормали. Таким образом, *в общем случае каждая падающая волна вызывает две преломленные волны*. Каждой преломленной волне соответствует направление луча и лучевая скорость, описывающие распространение энергии в кристалле *). Это явление носит название *двойного лучепреломления*. В качестве иллюстрации напомним хорошо известный эффект появления двух изображений при рассматривании небольшого объекта через пластинку исландского шпата.

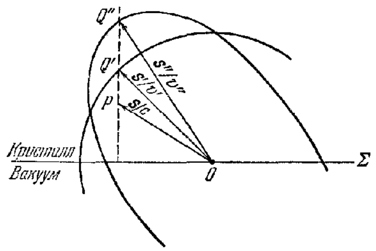


Рис. 14.10 К определению допустимых направлений волновых нормалей при двойном лучепреломлении.

Считая граничную плоскость Σ плоскостью $z=0$, перепишем (22) в виде равенств

$$\frac{xs_x+ys_y}{c}=\frac{xs'_x+ys'_y}{v'}=\frac{xs''_x+ys''_y}{v''}, \tag{23}$$

которые должны удовлетворяться при всех значениях x и y. Отсюда следует, во-первых, что $s'_x/s_x=s'_y/s_y$ и $s''_x/s_x=s''_y/s_y$, т. е. *оба преломленных луча лежат в плоскости падения*. Кроме того, если θ_i, θ'_t и θ''_t — углы, которые падающая и две прошедшие волны образуют с осью, то из (23) получим

$$\frac{\sin\theta_i}{\sin\theta'_t}=\frac{c}{v'},\qquad \frac{\sin\theta_i}{\sin\theta''_t}=\frac{c}{v''}. \tag{24}$$

Таким образом, каждая из прошедших в кристалл волн подчиняется такому же *закону преломления*, как и в случае изотропных сред. Однако скорость v здесь зависит от θ_t, так что определение направления распространения в кристалле становится более сложным В одноосном кристалле одна из оболочек обратной поверхности волновых нормалей является сферической и, значит, фазовая скорость *одной* из проходящих волн не зависит от θ_t. Это и есть *обыкновенная волна*.

Для специального случая нормального падения ($\theta_i=0$) имеем $\theta'_t=\theta''_t=0$ и, следовательно, обе волновые нормали в кристалле совпадают и направлены перпендикулярно к Σ. Другой специальный случай, представляющий большой теоретический интерес,— распространение волны в направлении одной из оптических осей двухосного кристалла. Возникающее при этом явление известно как коническая рефракция. Оно и рассматривается ниже.

б. Коническая рефракция. Выше мы отмечали, что когда в двухосном кристалле $\mathbf{s}$ совпадает с одной из оптических осей волновых нормалей, соотношение между $\mathbf{s}$ и $\mathbf{t}$ имеет особенность. Прежде чем обсуждать явления преломления при распространении волн в данном, специальном, направлении, необходимо исследовать природу этой особенности.

---

*) Возникновение двойного лучепреломления можно также пояснить, обобщая построение Гюйгенса (см п 3 3.3) на анизотропные среды. Однако точное рассмотрение такого подхода отнюдь не так просто, как это обычно излагается в учебниках При построении Гюйгенса оперируют с элементарными волнами, расходящимися из точечных источников, а не с независимыми плоскими волнами, законы распространения которых, как правило, принимаются без доказательства. Некоторые трудности, присущие такому подходу, обсуждались в работах [8, 9].

В п. 14.2.3 мы показали, что векторы электрического смещения **D**, соответствующие направлению волновой нормали **s**, параллельны главным осям перпендикулярного к **s** эллиптического сечения эллипсоида волновых нормалей. В специальном случае совпадения направления вектора **s** с направлением оптической оси волновых нормалей, это сечение становится круговым, так что допустимо любое направление **D**, перпендикулярное к **s**. Следовательно, в данном случае возможно бесконечное число направлений электрического вектора **E** (которые находят для каждого **D** из (14.1.1)) и лучевых векторов **t** (определяемых согласно рис. 14.1). Покажем, что все эти векторы **t** лежат на конической поверхности.

Пусть $\mathbf{s}'(s_x', 0, s_z')$ — единичный вектор, направленный вдоль одной из оптических осей волновых нормалей, причем главные диэлектрические оси выбраны в соответствии с неравенством (5). Тогда s_x' и s_z' связаны соотношением (11), и допустимые векторы **D**, будучи перпендикулярны к **s**′, удовлетворяют соотношению

$$s_x'D_x + s_z'D_z = 0.$$

Переходя к компонентам соответствующих векторов **E**, его можно переписать в виде

$$s_x'\varepsilon_x E_x + s_z'\varepsilon_z E_z = 0. \tag{25}$$

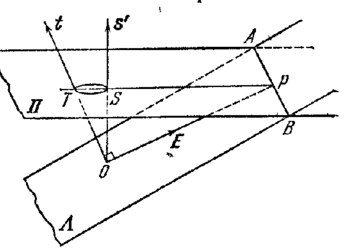


Рис. 14.11. К определению положения лучей, соответствующих оптической оси волновых нормалей в двухосном кристалле.

Пусть на рис. 14.11 П — любая плоскость, перпендикулярная к **s**′, и пусть прямая, на которой лежит вектор **E**, пересекает ее в точке P. Так как, согласно (25), все векторы **E** должны лежать на плоскости Λ, перпендикулярной к вектору с компонентами $(s_x'\varepsilon_x, 0, s_z'\varepsilon_z)$, то все возможные точки P должны лежать на прямой AB, по которой пересекаются плоскости П и Λ. Пусть лучевой вектор **t** компланарен с **E** и **s**′ и перпендикулярен к **E**. Пусть векторы **t** и **s**′ пересекают плоскость П соответственно в точках T и S. Тогда из подобия треугольников находим

$$TS \cdot SP = OS^2 = \text{const}. \tag{26}$$

Так как геометрическое место точек P — прямая линия AB, то геометрическим местом точек T является кривая, обратная AB, т. е. окружность *), проходящая через центр инверсии S, с касательной в точке S, параллельной AB. Следовательно, *оптической оси волновых нормалей соответствует бесконечное число лучей, которые образуют коническую поверхность.* Этот конус не круговой, так как центр круга не совпадает с основанием перпендикуляра из точки O на плоскость П.

Если векторы **E**, **s**′ и **t** лежат в плоскости xz, то направление **t** должно совпадать с направлением вектора $(s_x'\varepsilon_x, 0, s_z'\varepsilon_z)$, к которому **E** всегда перпендикулярен. Если оно образует угол φ с осью z, то угол раствора конуса χ в этой плоскости определяется выражением

$$\operatorname{tg}\chi = \operatorname{tg}(\beta - \varphi) = \frac{\dfrac{s_x'}{s_z'} - \dfrac{s_x'\varepsilon_x}{s_z'\varepsilon_z}}{1 + \dfrac{\varepsilon_x (s_x')^2}{\varepsilon_z (s_z')^2}} = \frac{1}{v_y^2}\sqrt{(v_x^2 - v_y^2)(v_y^2 - v_z^2)}, \tag{27}$$

где использовано соотношение (11). Обычно $(v_x^2 - v_y^2)/v_y^2 \ll 1$ и $(v_y^2 - v_z^2)/v_y^2 \ll 1$, так что получающийся конус очень мало отличается от кругового с углом χ.

*) См., например, [10].

Точно так же можно показать, что *лучевой оптической оси соответствует бесконечное число волновых нормалей, которые образуют коническую поверхность.* Угол раствора этого конуса ψ определяется соотношением, соответствующим (27), т. е.

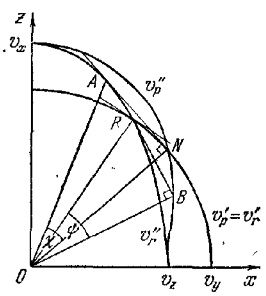


Рис. 14.12. Коническая рефракция: построение конусов.

$$\operatorname{tg}\psi=\frac{1}{v_xv_z}\sqrt{(v_x^2-v_y^2)(v_y^2-v_z^2)}=\frac{v_y^2}{v_xv_z}\operatorname{tg}\chi. \quad (28)$$

Полученные результаты удобно пояснить, прибегая к поверхности нормалей и лучевой поверхности и используя найденный в § 14.2.3 результат, указывающий, что поверхность нормалей представляет собой поверхность оснований перпендикуляров к лучевой поверхности. Сечение этих поверхностей плоскостью xz показано на рис. 14.12. Поверхность нормалей пересекает эту плоскость по окружности радиуса $v_p'=v_y$ и овалу с полярным радиусом v_p'', тогда как лучевая поверхность пересекает ее по той же окружности с $v_r'=v_y$ и эллипсу с v_r''. Если окружность и овал пересекаются в точке N, то линия ON совпадает с направлением оптической оси волновых нормалей, а плоскость, проведенная через N перпендикулярно к ON, должна касаться лучевой поверхности во всех точках, в которых конус допустимых лучевых направлений пересекает ее. Таким образом, лучевая поверхность обладает необычным свойством: некоторые касательные плоскости касаются ее в бесконечном числе точек *).

Лучевая оптическая ось совпадает с прямой OR, где R — точка пересечения двух оболочек лучевой поверхности; обе оболочки пересекаются таким образом, что в точке R имеется бесконечное число касательных плоскостей, нормали к которым образуют коническую поверхность. Нормали к этим плоскостям, проведенные из точки O, образуют конус направлений волновых нормалей, соответствующих направлению луча OR. Углы растворов χ и ψ обоих конусов также показаны на рис. 14.12. По причине, которая вскоре станет ясной, конус, принадлежащий к N, т. е. образованный такими лучами, как OA, называется *конусом внутренней конической рефракции.* Конус, принадлежащий к R, т. е. образованный такими волновыми нормалями, как OB, называется *конусом внешней конической рефракции.*

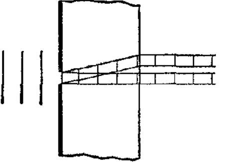


Рис. 14.13. Внутренняя коническая рефракция.

Возьмем пластинку двухосного кристалла, например арагонита, вырезанную так, что две ее параллельные грани перпендикулярны к оптической оси волновых нормалей. Если на такую пластинку нормально к одной из параллельных граней падает узкий пучок монохроматического света, то внутри пластинки энергия будет распространяться в полом конусе, конусе внутренней конической рефракции. При выходе с противоположной стороны световой пучок образует полый цилиндр (рис. 14.13). На экране, параллельном грани нашей кристаллической пластинки, следует ожидать появления яркого *круглого кольца.* Это замечательное явление было предсказано Вильямом Р. Гамильтоном в 1832 г., а через год его наблюдал Ллойд, исследовавший по предложению Гамильтона арагонит. Успех эксперимента послужил одним из наиболее четких подтверждений волновой теории света, развитой Френелем, и в очень сильной степени способствовал ее всеобщему признанию (см. «Историческое введение», стр. 17).

*) Свойства этой поверхности подробно рассмотрены в [11].

Практически демонстрация явления конической рефракции не так проста, как было описано выше, поскольку, конечно, невозможно получить строго параллельный пучок монохроматического света. В эксперименте нам всегда приходится использовать пучки с конечной угловой апертурой. В таком случае, как впервые показали Поггендорф [12] и Хайдингер [13], будут наблюдаться два ярких кольца, разделенных тонким темным кольцом (рис. 14.14). В первых экспериментах Ллойда эту структуру наблюдать не удалось, поскольку отверстия, ограничивающие ширину используемого им пучка, были слишком велики, и оба ярких кольца сливались в одно. Возникающую структуру не могли объяснить в течение долгого времени после ее открытия, пока Фохт [14] не предложил интерпретацию, которая в конечном счете сводится к следующему.

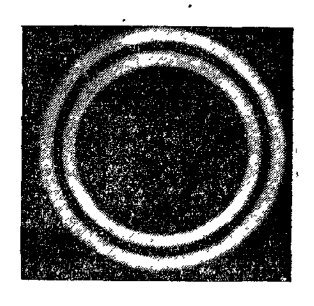


Рис. 14.14. Распределение света, возникающее вследствие конической рефракции.

Мы должны рассмотреть распространение волн, нормали которых слегка наклонены к оптической оси. Каждой из волновых нормалей соответствуют два луча внутри кристалла, и следует ожидать, что их направления мало отличаются от направлений образующих конуса внутренней конической рефракции. Чтобы найти распределение прошедших лучей, необходимо рассмотреть часть лучевой поверхности вблизи окружности, по которой она касается плоскости AN (см. рис. 14.12). Эта часть поверхности напоминает часть надутой автомобильной камеры, а касательная плоскость — плоскую доску, лежащую на ней. На рис. 14.15 показано сечение этой части поверхности плоскостью xz. Две точки на лучевой поверхности, которые соответствуют направлениям двух лучей, относящихся к данному направлению волновой нормали $\mathbf{s}$, определяются как точки касания этой поверхности двумя плоскостями, перпендикулярными к $\mathbf{s}$ (см. рис. 14.15). Когда волновая нормаль ON слегка отклоняется от оптической оси, вместо одной касательной плоскости возникают две параллельные друг другу плоскости, одна из них при этом перемещается над лучевой поверхностью, причем точка ее касания движется от центра касательной окружности к точке F. Другая плоскость (ее невозможно показать на нашей модели, потому что она должна пересекать нашу камеру) перемещается так, что точка ее касания движется по направлению к точке G. Рис. 14.15 иллюстрирует это для смещения волновой нормали в плоскости xz, но та же картина будет наблюдаться и при смещении в любом другом направлении.

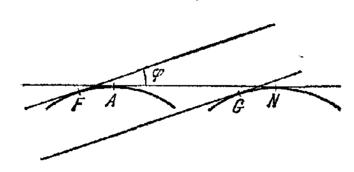


Рис. 14.15. К определению положения лучей, относящихся к тем волновым нормалям, которые составляют небольшие углы с оптической осью волновых нормалей.

Из изложенных выше рассуждений следует, что каждая падающая волна, нормаль к которой образует небольшой угол φ с оптической осью, приведет к появлению пары лучей, составляющих углы $^1/_2\chi + a\varphi$ и $^1/_2\chi - a\varphi$ с осью конуса внутренней конической рефракции, где a — некоторая постоянная. Таким образом, вся энергия, сосредоточенная в падающем пучке в интервале углов от φ до $\varphi + d\varphi$, окажется в двух конусах, половина угла раствора которых равна $^1/_2\chi \pm a\varphi$, а угловое расстояние между ними составляет $a d\varphi$. Но соответствующая энергия в падающем пучке пропорциональна $\varphi d\varphi$. Отсюда следует, что при углах $^1/_2\chi \pm a\varphi$ интенсивность в конусе лучей пропорциональна φ и, в частности, при $\varphi = 0$ она равна нулю. Таким образом, нужно

ожидать появления двух ярких колец с темным промежутком между ними, что и наблюдается.

Внешняя коническая рефракция демонстрирует уже установленный ранее факт существования целого конуса направлений волновых нормалей, соответствующего данному направлению луча. Это явление наблюдалось на кристаллической пластинке, вырезанной так, что ее грани перпендикулярны к лучевой оптической оси. На гранях пластинки точно друг против друга находятся два небольших отверстия, одно из которых освещается сходящимся пучком света, как показано на рис. 14.16. Второго отверстия будут достигать лишь те лучи, направления которых очень близки к направлению лучевой оптической оси, так что нормали всех волн, достигающих второго отверстия, располагаются вблизи конуса внешней конической рефракции. Поэтому из кристалла будет выходить конус света. Из-за преломления при выходе из кристалла угол раствора этого конуса будет больше истинного угла ψ внешней конической рефракции. И в данном случае на экране, параллельном грани кристалла, наблюдаются два концентрических кольца света, причем объяснение этого подобно объяснению, данному в случае внутренней конической рефракции.

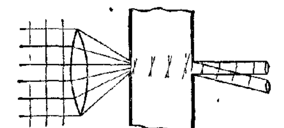


Рис. 14.16. Внешняя коническая рефракция.

§ 14.4. Измерения в кристаллооптике

В настоящем разделе мы кратко опишем методы определения свойств кристалла (т. е. является ли он одноосным или двухосным), положения его оптических осей и значений его главных показателей преломления. Как мы увидим, оптические оси можно определить при наблюдении интерференционных полос на кристаллических пластинках; характер интерференционной картины ясно указывает на взаимное расположение оптических осей и граней пластинки. Главные показатели преломления можно найти с помощью кристаллических призм, измеряя углы отклонения или полного внутреннего отражения.

Предварительно необходимо рассмотреть методы получения и анализа поляризованного света.

Рис. 14.17. Призма Николя.

14.4.1. Призма Николя. Одним из наиболее широко используемых приспособлений для получения линейно поляризованного света является *призма Николя* [15]. Она состоит из природного кристалла исландского шпата в виде ромбоэдра (одноосный кристалл), разрезанного на две равные части вдоль диагональной плоскости (на рис. 14.17 она показана линией AC). Эти части склеены вместе канадским бальзамом. Длина ромба в три раза больше его ширины, а углы при B и D его главного сечения равны 71°; торцевые грани AD и BC сполированы так, чтобы уменьшить этот угол до 68°.

Луч света, распространяющийся в направлении L параллельно длинной стороне, разделяется на два луча, обыкновенный и необыкновенный. Для первого луча канадский бальзам является оптически менее плотной средой, для второго — более плотной (для кристалла исландского шпата $n_o = 1{,}66$, $n_e = 1{,}49$; для канадского бальзама $n = 1{,}53$). Легко проверить по формулам, приведенным в п. 1.5.4, что на границах с канадским бальзамом для обыкновенного луча выполняются условия полного внутреннего отражения. Этот луч полностью отражается в направлении к зачерненной грани DC, где он и поглощается. Необыкновенный луч проходит через призму, практически не испыты-

вая бокового смещения, и выходит из нее линейно поляризованным, причем его вектор **D** расположен в плоскости главного сечения (см. п. 14.3.2). Таким образом, *призма Николя дает линейно поляризованный свет с известным направлением колебаний.*

Если падающий луч наклонен к краю ромба, призма Николя все же будет служить поляризатором при условии, что для обыкновенного луча угол падения на слой канадского бальзама не меньше критического. В воздухе призма эффективно работает, если угловая апертура падающих лучей не превышает *) 30°.

Весьма полную линейную поляризацию света можно получить, пропуская естественный свет через пластинку поглощающего кристаллического вещества с существенно различными коэффициентами поглощения для двух направлений колебаний. Мы рассмотрим такие «поляроиды» в п. 14.6.3.

Устройство, аналогичное призме Николя, которое превращает свет с любыми типами поляризации в линейно поляризованный свет, называется *поляризатором*. Его можно также использовать в качестве *анализатора*, т. е. детектора линейно поляризованного света и его направления колебаний. Для детектирования линейно поляризованного света с помощью призмы Николя нужно лишь поворачивать призму вокруг ее продольной оси и установить, существует ли положение, когда свет через нее не проходит. Если такое положение существует, то свет линейно поляризован и направление колебаний его вектора **D** перпендикулярно к главному сечению призмы.

14.4.2. Компенсаторы. Кристаллическое вещество можно также использовать для исследования эллиптически поляризованного света, т. е. для определения направления осей его эллипса поляризации и отношения их длин. Для этой цели применяется кристаллическая пластинка из подходящего материала нужной толщины. С ее помощью вводится разность фаз между колебаниями в двух взаимно ортогональных направлениях. В частности, можно ввести такую разность фаз, чтобы преобразовать эллиптически поляризованный свет в линейно поляризованный. Требующаяся информация об эллиптически поляризованном свете получается тогда из анализа этого линейно поляризованного света. Такой прибор называется *компенсатором*, так как его функция состоит в компенсации разности фаз **).

а. Четвертьволновая пластинка. Рассмотрим плоскопараллельную кристаллическую пластинку толщиной h. Пусть ось z перпендикулярна к пластинке, а направления осей x и y совпадают с направлениями соответствующих колебаний вектора **D**. Предположим, что пластинка поворачивается вокруг нормали до тех пор, пока оси x и y не станут параллельными главным осям поляризации падающего света. Тогда компоненты вектора **D** для падающего света можно записать в виде

$$D_x^{(i)} = a\cos\omega t, \qquad D_y^{(i)} = b\sin\omega t. \tag{1}$$

Поскольку скорости обоих лучей различны, изменения фаз компонент при прохождении через пластинку будут неодинаковыми. Пренебрегая потерями на отражение, можно записать компоненты вектора **D** для света, выходящего из пластинки, в виде

$$D_x^{(t)} = a\cos(\omega t + \delta'), \qquad D_y^{(t)} = b\sin(\omega t + \delta''), \tag{2}$$

где

$$\delta' = \frac{2\pi}{\lambda}n'h, \qquad \delta'' = \frac{2\pi}{\lambda}n''h, \tag{3}$$

*) Существует несколько модификаций призмы Николя, которые могут работать с еще бо́льшими апертурами. Их описание см., например, в [16] или [17].

**) Подробный обзор, посвященный компенсаторам, приведен в работе [18]. Обзор методов, применяемых для анализа поляризованного света, дан в статье [19].

Аппаратом для теоретического анализа полностью или частично поляризованного света служат матрицы когерентности или параметры Стокса, рассмотренные в § 10.8.

а λ — длина волны в вакууме *). Следовательно, вносимая пластинкой разность фаз равна

$$\delta''-\delta'=\frac{2\pi}{\lambda}(n''-n')h. \qquad (4)$$

В частности, если прошедший свет *линейно поляризован*, мы должны получить $\delta''-\delta'=\pm\pi/2$ или в более общем случае $\delta''-\delta'=(2m+1)\pi/2$, где m — любое целое число При этом пластинка должна иметь толщину

$$h=\left|\frac{2m+1}{n''-n'}\right|\frac{\lambda}{4}. \qquad (5)$$

При наличии такого компенсатора направление колебаний в прошедшем линейно поляризованном свете определяется соотношением

$$\frac{D_y^{(t)}}{D_x^{(t)}}=\pm\frac{b}{a}. \qquad (6)$$

Компенсатор, вносящий разность фаз $|\delta''-\delta'|=\pi/2$, т. е. такой, для которого разность двух оптических толщин составляет четверть волны, называется *четвертьволновой пластинкой* [20]. Ее легко изготовить из листочка слюды (двухосный кристалл), расщепленной до толщины $h=\lambda/(4|n''-n'|)$.

С помощью такой пластинки можно следующим образом анализировать эллиптически поляризованный свет пучок света пропускают через четвертьволновую пластинку, а затем через призму Николя, и оба кристалла независимо друг от друга поворачивают до тех пор, пока поле, наблюдаемое через призму Николя, не станет совсем темным. В этом положении оси слюдяной пластинки параллельны осям эллипса поляризации падающего света, а призма Николя, согласно (6), расположена так, что она гасит линейно поляризованный свет, вектор **D** которого составляет с осью x угол $\operatorname{arctg}(b/a)$. Таким образом, определяют ориентацию осей эллипса и их отношение.

Если падающий свет поляризован по кругу, то $b=+a$ или $b=-a$, а прошедший свет линейно поляризован, причем направление вектора **D** составляет с осью x соответственно углы 45° или 135°. Первый из углов соответствует левой, второй — правой поляризации.

Так как толщина четвертьволновой пластинки зависит от λ, точная компенсация возможна только для монохроматического света одной определенной длины волны. Для достижения компенсации в случае света любой заданной длины волны следует использовать не одну плоскопараллельную пластинку, а клин или комбинацию клиньев Рассмотрим некоторые компенсаторы такого типа.

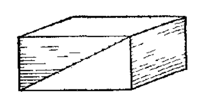


Рис 14.18. Компенсатор Бабине.

б *Компенсатор Бабине*. Компенсатор, разработанный Бабине [21], позволяет получить любые разности фаз в том числе и нулевые. Он состоит из двух кварцевых клиньев (положительный одноосный кристалл) с одинаковыми острыми углами. Клинья расположены так, как показано на рис 14 18, и могут смещаться вдоль плоскости соприкосновения, образуя, таким образом, параллельную пластинку варьируемой толщины. В одном из клиньев оптическая ось параллельна, в другом — перпендикулярна ребру.

Пусть n_o и n_e — показатели преломления кварца для обыкновенного и необыкновенного лучей, а h_1 и h_2 — толщины двух клиньев в некоторой определенной точке После прохождения обоих клиньев разность фаз между лучами становится равной

$$\delta=\frac{2\pi}{\lambda}(n_e-n_o)(h_1-h_2). \qquad (7)$$

*) В отличие от § 1.3 здесь мы пишем λ, а не λ_0.

Легко понять, почему в это выражение входит разность толщин. В самом деле, скорость луча, для которого направление колебаний вектора $\mathbf{D}$ перпендикулярно к главной оси (обыкновенный луч), превышает скорость необыкновенного луча. Поэтому компонента, в которой направление колебаний перпендикулярно к ребру клина, будет опережать вторую компоненту в одном клине и отставать в другом. Толщины h_1 и h_2, а следовательно, и разность фаз δ непрерывно меняются, если точка падения света движется вдоль пластинки, причем δ равно нулю в середине. В результате получится серия линий, где прошедший свет будет линейно поляризован и, значит, его можно погасить путем соответствующей ориентации призмы Николя.

Если выбрать ось x параллельно ребру клина, то эллиптическую поляризацию падающего света можно представить следующим образом:

$$D_x^{(i)} = a_1 \cos(\omega t + \delta_0'), \qquad D_y^{(i)} = a_2 \cos(\omega t + \delta_0''). \tag{8}$$

Поляризация света, прошедшего через компенсатор, записывается в виде

$$D_x^{(t)} = a_1 \cos(\omega t + \delta'), \qquad D_y^{(t)} = a_2 \cos(\omega t + \delta''), \tag{9}$$

где

$$\delta' = \delta_0' + \frac{2\pi}{\lambda}(n_o h_1 + n_e h_2), \qquad \delta'' = \delta_0'' + \frac{2\pi}{\lambda}(n_e h_1 + n_o h_2). \tag{10}$$

Для линейной поляризации выходящего света необходимо, чтобы

$$\delta'' - \delta' = m\pi \ (m = 0, \ \pm 1, \ \pm 2, \ \ldots),$$

т. е.

$$\delta_0'' - \delta_0' = -\frac{2\pi}{\lambda}(n_e - n_o)(h_1 - h_2) + m\pi. \tag{11}$$

Направление колебаний при этом определяется соотношением

$$\frac{D_y^{(t)}}{D_x^{(t)}} = \pm \frac{a_2}{a_1}. \tag{12}$$

Предположим вначале, что призма Николя помещена перед компенсатором, так что на компенсатор падает линейно поляризованный свет. Тогда $\delta_0'' = \delta_0'$, и если за компенсатором следует анализатор, скрещенный с поляризующей призмой Николя, то появятся темные полосы, идущие параллельно ребру клина через точки, для которых правая часть уравнения (11) кратна π. Эти темные полосы определяют *нулевое положение*. Если теперь исследовать эллиптически поляризованный свет только посредством компенсатора и анализатора, то смещение темных полос от нулевого положения сразу же определит разность фаз $\delta_0'' - \delta_0'$ для падающего света. Величину отношения компонент $\mathbf{D}^{(i)}$, параллельной и перпендикулярной ребру, можно найти из ориентации анализатора, используя (12). Эти данные позволяют с помощью формул, приведенных в п. 1.4.2, определить положение главных осей эллипса поляризации и их отношение.

в. Компенсатор Солейля. В некоторых случаях необходимо создавать разность фаз (положительную, отрицательную или нулевую), остающуюся постоянной во всем поле зрения. Этого можно достичь с помощью компенсатора, разработанного Солейлем [22]. Он состоит из двух кварцевых клиньев A и A', которые, как и в случае компенсатора Бабине, образуют плоскопараллельную пластинку, но с тем отличием, что оптические оси *обоих* клиньев параллельны ребрам. Нижний клин склеен с плоскопараллельной кварцевой пластинкой B, оптическая ось которой перпендикулярна к ребру клина (рис. 14.19). Эффективная оптическая разность хода, которую вносит этот компенсатор между двумя лучами, очевидно, равна $(n_e - n_o)[h_B - (h_A + h_{A'})]$. В нулевом положении, когда эта разность исчезает, все поле зрения можно сделать темным

при помощи соответствующим образом ориентированного анализатора. Эффективную оптическую разность хода можно изменять, смещая верхний клин, но для каждого положения разность хода остается постоянной по всему полю. Анализ эллиптически поляризованного света проводится примерно так же, как и с помощью компенсатора Бабине.

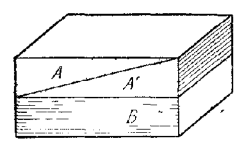


Рис. 14.19. Компенсатор Солейля.

г. Компенсатор Берека. Берек [23] разработал компенсатор, используемый в биологической микроскопии (например, для измерения разностей хода в нитях двоякопреломляющих материалов). Он состоит из пластины одноосного кристаллического вещества с оптической осью, перпендикулярной к граням пластины. Разность фаз между обыкновенным и необыкновенным лучами вводится путем наклона пластины. Ее монтируют так, что она может поворачиваться с помощью проградуированного диска, который калибруется прямо в единицах вносимой разности хода.

14.4.3. Интерференция в кристаллических пластинках. Как отмечалось, выше, направления оптических осей можно определить путем наблюдения интерференционных явлений в кристаллических пластинках. Возникающие при этом интерференционные эффекты, которые служат поразительной демонстрацией согласия теории и эксперимента в кристаллооптике, вполне заслуживают самостоятельного изучения.

Рассмотрим вначале пучок линейно поляризованного света, выходящий из поляризатора и падающий нормально на плоскопараллельную кристаллическую пластинку толщиной h. Входя в пластинку, каждый луч делится на два луча, распространяющиеся с различной скоростью, причем их векторы **D** колеблются в двух взаимно ортогональных направлениях, перпендикулярных направлению нормали к пластинке. Лучи выходят из пластинки с определенной разностью фаз δ. Если анализирующая призма Николя расположена за пластинкой, то она пропускает определенные компоненты этих двух колебаний, которые могут интерферировать в фокальной плоскости линзы, расположенной за анализатором.

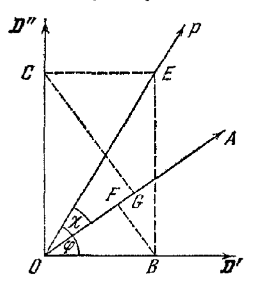


Рис 14.20 К определению компонент колебаний, пропускаемых поляризатором и анализатором.

На рис. 14.20 плоскость чертежа параллельна пластинке. **D**$'$ и **D**$''$ — два взаимно ортогональных направления колебаний в кристалле, а OP и OA — направления колебаний, которые пропускаются соответственно поляризатором и анализатором. Пусть φ — угол, который OP образует с **D**$'$, а χ — угол между OA и OP. Амплитуда световой волны, падающей на пластинку, представляется вектором OE (он параллелен OP); его составляющие в направлениях **D**$'$ и **D**$''$ равны

$$OB = E\cos\varphi, \qquad OC = E\sin\varphi. \tag{13}$$

Анализатор пропускает лишь составляющие, параллельные OA, амплитуды которых, как легко получить из рисунка и из выражения (13), равны

$$OF = E\cos\varphi\cos(\varphi-\chi), \qquad OG = E\sin\varphi\sin(\varphi-\chi). \tag{14}$$

Покидая пластинку, компоненты различаются по фазе на величину

$$\delta = \frac{2\pi}{\lambda}(n''-n')h. \tag{15}$$

Согласно (7.2.15) интенсивность, получающаяся в результате интерференции

двух монохроматических волн с разностью фаз δ, равна

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\delta,$$

где I_1 и I_2 — интенсивности (квадраты амплитуд) обеих волн. Если амплитуды имеют вид (14), то

$$I = E^2 \left\{\cos^2\chi - \sin 2\varphi \sin 2(\varphi - \chi) \sin^2\frac{\delta}{2}\right\}, \tag{16}$$

где использовано тождество $\cos\delta = 1 - 2\sin^2(\delta/2)$.

Если удалить пластинку ($\delta = 0$), то интенсивность будет равна $I = E^2\cos^2\chi$, следовательно, второй член в (16) представляет собой влияние кристаллической пластинки.

Рассмотрим два важных специальных случая.

а. *Анализатор и поляризатор параллельны* ($\chi = 0$). В этом случае (16) примет вид

$$I_{\parallel} = E^2\left(1 - \sin^2 2\varphi \sin^2\frac{\delta}{2}\right). \tag{17}$$

Имеется *максимум пропускания* при

$$\varphi = 0, \quad \frac{\pi}{2}, \ \pi, \ \ldots, \tag{18}$$

т. е. *когда направление колебаний, пропущенных анализатором, совпадает с одним из направлений колебаний в пластинке.* Между положениями, определяемыми (18), располагаются *минимумы пропускания*, которые задаются соотношением $\sin 2\varphi = \pm 1$, т. е.

$$\varphi = \frac{\pi}{4}, \ \frac{3\pi}{4}, \ \frac{5\pi}{4}, \ \ldots, \tag{19}$$

причем минимумы интенсивности равны

$$I_{\parallel\,\text{мин}} = E^2\left(1 - \sin^2\frac{\delta}{2}\right) = E^2\cos^2\frac{\delta}{2}. \tag{20}$$

В минимуме нет полной темноты, если δ не равно нечетному числу π, т. е. если толщина пластинки не равна одному из ряда специальных значений, зависящих от длины волны света.

б. *Анализатор и поляризатор взаимно перпендикулярны* ($\chi = \pi/2$). В данном случае из (16) следует

$$I_{\perp} = E^2 \sin^2 2\varphi \sin^2\frac{\delta}{2}. \tag{21}$$

Сравнение с (17) показывает, что возникающие здесь интерференционные явления оказываются *дополнительными* к случаю *а*. Имеются *минимумы, соответствующие полной темноте*, когда φ принимает одно из значений (18), и относительные *максимумы* для промежуточных положений (19), равные

$$I_{\perp\,\text{макс}} = E^2\sin^2\frac{\delta}{2}. \tag{22}$$

Подобные явления можно использовать для получения строго дополнительных цветов. Для этого необходимо лишь пропустить пучок белого света через две призмы Николя, разделенные кристаллической пластинкой, причем призмы должны быть вначале параллельны друг другу ($\chi = 0$), а затем скрещены ($\chi = \pi/2$). Для создания однородного поля существенно иметь достаточно параллельный пучок, чтобы не вносить заметной разности фаз.

Если свет, падающий на первый поляризатор, испускается протяженным некогерентным источником, расположенным в фокальной плоскости линзы, то каждая точка этого источника независимо от остальных его точек будет создавать в сопряженной точке некоторую интенсивность. Наблюдающееся в плоскости изображения распределение света, которое можно описать с

помощью *кривых равной интенсивности*, обусловливает так называемую *интерференционную картину* кристаллической пластинки. Каждой точке в плоскости изображения соответствует направление параллельных лучей, входящих и выходящих из кристалла, и поэтому необходимо рассмотреть изменение разности фаз δ с изменением этого направления. Достаточно рассмотреть лишь случай скрещенных поляризатора и анализатора ($\chi = \pi/2$), так как при их параллельности картина будет дополнительной, другие случаи приводят к менее четким интерференционным картинам.

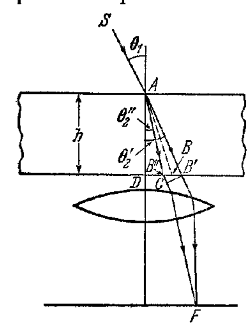


Рис. 14.21. К определению разности фаз, приобретаемой двумя волнами, прошедшими через кристаллическую пластинку.

Интенсивность, соответствующая заданному направлению падения, зависит от φ и δ. Полезно рассмотреть влияние изменения каждой из этих величин по отдельности. Кривые, вдоль которых φ постоянно, называются *изогирами*, кривые, вдоль которых δ постоянно,—*изохроматами*. Изогиры зависят от ориентации оптических осей в пластинке и не зависят от толщины пластинки и длины волны, если отсутствует дисперсия осей. Изохроматы зависят от направления волновых нормалей и от толщины пластинки и называются так потому, что при использовании белого света они представляют собой линии одинакового цвета. Принадлежащие этим семействам кривые, вдоль которых интенсивность равна нулю, называются *главными изогирами* и *главными изохроматами*; согласно (21) они задаются соотношениями $\sin 2\varphi = 0$ и $\sin \frac{1}{2}\delta = 0$. Соответствующее таким кривым состояние поляризации света совпадает с состоянием поляризации света, падающего на кристалл. Это связано с тем, что для главных изогир направление колебаний, пропускаемых анализатором, совпадает с одним из направлений колебаний в кристалле, а для главных изохромат разность фаз между двумя выходящими пучками кратна 2π. Обе системы кривых налагаются друг на друга, но их можно изучать по отдельности.

Прежде чем перейти к исследованию формы этих кривых, необходимо рассмотреть зависимость разности фаз δ от угла падения. Пусть SA, AB', AB'' — волновые нормали к падающей и к двум преломленным волнам в точке A, а θ_1, θ_2', θ_2'' — соответственно угол падения и два угла преломления (рис. 14.21). Далее пусть λ — длина волны в первой среде (воздух) и $\lambda' = \lambda/n'$, $\lambda'' = \lambda/n''$ — длины обеих преломленных волн. Лучи из пластинки будут выходить параллельно друг другу и волновой нормали к падающей волне (см. рис. 14.21) с разностью фаз

$$\delta = 2\pi \left[\frac{AB''}{\lambda''} + \frac{B''C}{\lambda} - \frac{AB'}{\lambda'}\right], \tag{23}$$

где

$$AB' = \frac{h}{\cos\theta_2'}, \qquad AB'' = \frac{h}{\cos\theta_2''} \tag{24}$$

и

$$B'C = B''B' \sin\theta_1 = h \sin\theta_1 (\operatorname{tg}\theta_2' - \operatorname{tg}\theta_2''). \tag{25}$$

Подставляя последние два выражения в (23), получим

$$\delta = 2\pi h \left[\frac{1}{\cos\theta_2''}\left(\frac{1}{\lambda''} - \frac{\sin\theta_1 \sin\theta_2''}{\lambda}\right) - \frac{1}{\cos\theta_2'}\left(\frac{1}{\lambda'} - \frac{\sin\theta_1 \sin\theta_2'}{\lambda}\right)\right]. \tag{26}$$

Используя закон преломления, мы можем заменить $\sin\theta_1/\lambda$ на $\sin\theta_2''/\lambda''$ в первых скобках и на $\sin\theta_2'/\lambda'$ во вторых скобках, и тогда получим

$$\delta = 2\pi h \left[\frac{\cos\theta_2''}{\lambda''} - \frac{\cos\theta_2'}{\lambda'}\right] = \frac{2\pi h}{\lambda}(n'' \cos\theta_2'' - n' \cos\theta_2'). \tag{27}$$

Так как разность $n'' - n'$ всегда мала по сравнению с n' и n'', можно заменить (27) приближенным выражением. В первом приближении имеем

$$n'' \cos\theta_2'' - n' \cos\theta_2' = (n'' - n') \frac{d}{dn}(n \cos\theta_2) = (n'' - n') \left[\cos\theta_2 - n \sin\theta_2 \frac{d\theta_2}{dn}\right], \quad (28)$$

где n — среднее значение n' и n'', θ_2 — соответствующее среднее углов θ_2' и θ_2''. Дифференцируя выражение для закона преломления $\sin\theta_1 = n \sin\theta_2$ при фиксированном значении θ_1, получим также

$$0 = \sin\theta_2 + n \cos\theta_2 \frac{d\theta_2}{dn}. \quad (29)$$

Следовательно, (28) можно записать в виде

$$n'' \cos\theta_2'' - n' \cos\theta_2' = \frac{1}{\cos\theta_2}(n'' - n'), \quad (30)$$

а подставив в (27) выражение (30), найдем

$$\delta = \frac{2\pi h}{\lambda \cos\theta_2}(n'' - n'). \quad (31)$$

Величина $h/\cos\theta_2$ представляет средний геометрический путь двух лучей в пластинке, а после умножения на $n'' - n'$ — соответствующую оптическую разность хода.

Возвращаясь к случаю протяженного источника, мы должны рассмотреть прохождение волн с различными направлениями распространения. Предположим, что эти направления образуют небольшие углы с нормалью к пластинке. Зададим каждую из падающих волн ее волновой нормалью в фиксированной точке A (см. рис. 14.21). Между точками F в фокальной плоскости линзы, где собираются волны, и точками B, где нормали проходящих волн AB пересекают нижнюю поверхность пластинки, существует однозначное соответствие (AB — это среднее из AB' и AB''). Так как наклон AB к нормали пластинки AD мал, точки F образуют слегка искаженное изображение — проекцию точек B. Следовательно, форма изохромат существенно зависит от местоположения точек B, для которых постоянна величина δ. В частности, для главных изохромат эта постоянная равна целому кратному 2π. Для исследования влияния изменения толщины пластинки мы должны лишь сместить плоскость, содержащую точки B, параллельно самой себе.

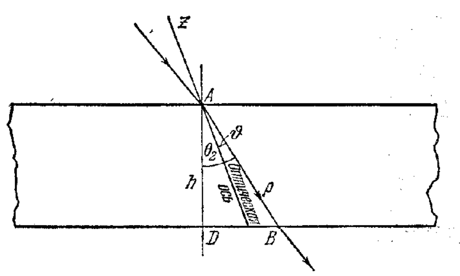


Рис. 14.22. К теории интерференционных картин, получающихся с кристаллическими пластинками.

Очевидно, можно получить все изохроматы, строя вокруг некоторой точки A *поверхности постоянной разности фаз* $\delta(h, \theta_2) = \text{const}$, называемые также *изохроматическими* поверхностями, и находя их пересечения с плоскостями $h = \text{const}$. При определении этих поверхностей необходимо помнить, что показатели преломления n' и n'', входящие в выражение (31) для δ, также зависят от θ_2.

Рассмотрим по отдельности форму интерференционных картин, полученных с пластинками из одноосного и двухосного кристаллов. Для задания положения точек B удобно воспользоваться полярным радиусом

$$\rho = AB = \frac{h}{\cos\theta_2} \quad (32)$$

и углом ϑ (или углами ϑ_1 и ϑ_2), который AB образует с направлением оптической оси (или осей) волновых нормалей кристаллической среды (рис. 14.22).

14.4.4. Интерференционные картины, получающиеся с пластинками одноосных кристаллов. В одноосном кристалле фазовые скорости, соответствующие направлению волновой нормали, образующему угол ϑ с оптической осью, связаны между собой, согласно (14 3.4), соотношением

$$(v'_p)^2-(v''_p)^2=(v_o^2-v_e^2)\sin^2\vartheta. \qquad (33)$$

Так как $v_p=c/n$ и аналогично для других скоростей, мы вправе написать

$$\frac{1}{(n')^2}-\frac{1}{(n'')^2}=\left(\frac{1}{n_o^2}-\frac{1}{n_e^2}\right)\sin^2\vartheta. \qquad (34)$$

Разность этих двух показателей преломления обычно мала по сравнению с их величинами, и поэтому последнее выражение приближенно можно представить в виде

$$n''-n'=(n_e-n_o)\sin^2\vartheta. \qquad (35)$$

Подставляя выражение (35) в (31) и используя (32), получим

$$\delta=\frac{2\pi\rho}{\lambda}(n_e-n_o)\sin^2\vartheta \qquad (36)$$

Следовательно, поверхности постоянной разности фаз определяются соотношением

$$\rho\sin^2\vartheta=C \qquad (C=\text{const}). \qquad (37)$$

Чтобы представить себе форму этих поверхностей, обратимся к декартовым осям координат с осью z, направленной вдоль оптической оси. Тогда

$$\left.\begin{aligned}\rho^2&=x^2+y^2+z^2,\\ \rho^2\sin^2\vartheta&=x^2+y^2,\end{aligned}\right\} \qquad (38)$$

и в соответствии с (37) поверхности постоянной разности фаз будут определяться уравнением

$$(x^2+y^2)^2=C^2(x^2+y^2+z^2). \qquad (39)$$

Эти поверхности можно получить при вращении кривых

$$x^4=C^2(x^2+z^2) \qquad (40)$$

вокруг оси z (типичная кривая такого вида показана на рис. 14 23). На больших расстояниях от начала координат ($z^2\gg x^2$) кривые асимптотически приближаются к параболам

$$x^2=\pm Cz. \qquad (41)$$

Вблизи оси x имеем

$$C^2=\frac{x^4}{x^2+z^2}=\frac{x^2}{1+z^2/x^2}=x^2\left(1-\frac{z^2}{x^2}+\ldots\right), \qquad (42)$$

так что кривая там аппроксимируется гиперболой

$$x^2-z^2=C^2. \qquad (43)$$

На рис. 14.23 эти параболы и гиперболы показаны пунктиром.

Теперь мы можем определить все изохроматы, просто беря сечения поверхности (39) плоскостями, находящимися на различном расстоянии h от начала координат, причем все возникающие изохроматы можно наблюдать на внешней грани кристаллической пластинки. Из рассмотрения рис. 14.23 очевидно, что формы изохромат зависят от ориентации этой грани пластинки относительно оптической оси. Если внешняя грань пластинки перпендикулярна к оптической оси, то ясно, что изохроматы имеют вид окружностей; если нормаль к грани образует небольшой угол с оптической осью, кривые сжимаются и переходят в эллипсы, если же эта нормаль образует большой угол с оптической осью, кривые приближаются к гиперболам.

Главные изогиры (кривые, для которых $\sin 2\varphi = 0$) имеют вид темного расплывчатого креста, образующие которого параллельны направлениям поляризатора и анализатора, а центр соответствует направлению волновой нормали, параллельному оптической оси. Это следует из того, что, как известно, при любом заданном направлении волновой нормали в кристалле направления колебаний параллельны и перпендикулярны главной плоскости, в которой лежат волновая нормаль и оптическая ось. Таким образом, всем направлениям, для которых главная плоскость параллельна направлению колебаний, пропускаемых либо поляризатором, либо анализатором, соответствуют в поле зрения темные области.

На рис. 14.24 показана типичная интерференционная картина, полученная с одноосным кристаллом; на ней ясно видны главные изогиры и изохроматы.

Интерференционные картины в плоскопараллельных кристаллических

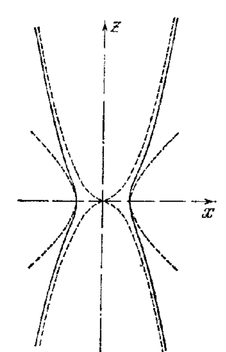


Рис. 14.23. Меридиональная кривая поверхности постоянной разности фаз для оптически одноосного кристалла

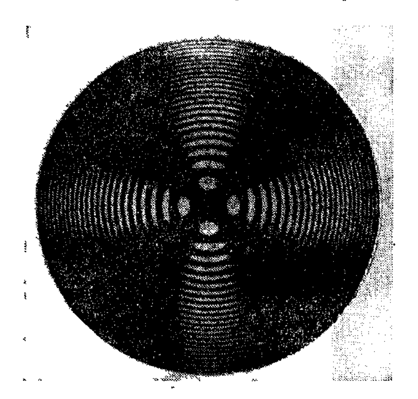


Рис. 14.24. Интерференционная картина, получаемая с плавиковым шпатом, вырезанным перпендикулярно оптической оси.
Призмы Николя скрещены

пластинках находят важное практическое применение в *полярископах*. Эти приборы используются для определения небольших количеств поляризованного света в пучке, который в основном неполяризован. В качестве примера приведем пластинку Савара [24]. Она состоит из двух кварцевых пластинок, каждая из которых вырезана таким образом, что ее оптическая ось составляет 45° с нормалью к пластинке. Обе пластинки склеены так, что плоскости, содержащие нормали к пластинкам и оптические оси, перпендикулярны друг к другу. Следовательно, направление колебаний, соответствующее обыкновенному лучу в одной пластинке, совпадает с направлением колебаний, соответствующим необыкновенному лучу в другой, а разности фаз имеют противоположные знаки. При нормальном прохождении волн через пластинки разности фаз точно компенсируют друг друга. Для случая почти нормального падения в таком устройстве разность фаз почти не меняется при изменении длины волны, что позволяет получать полосы в белом свете. Если между скрещенными призмами Николя поместить пластинку Савара так, чтобы ее главные плоскости составляли угол 45° с направлениями призм, то интерференционная картина будет состоять из почти прямых светлых и темных полос. Если убрать поляризующую призму и пропускать через систему частично поляризованный свет, то поляризованная составляющая будет создавать полосы, которые накладываются на однородный фон, обусловленный неполяризован-

ной компонентой. Эти полосы можно обнаружить даже при очень малой доле поляризованного света Полосы наиболее контрастны, когда плоскость колебаний падающего света составляет угол 45° с главными плоскостями пластинки Савара.

14.4.5. Интерференционные картины, получающиеся с пластинками двухосных кристаллов. Для плоскопараллельной пластинки двухосного кристалла мы получим вместо (33) более общее соотношение

$$(v_p')^2-(v_p'')^2=(v_x^2-v_z^2)\sin\vartheta_1\sin\vartheta_2, \tag{44}$$

которое следует из (14.3.18). Здесь ϑ_1 и ϑ_2 — углы между направлением волновой нормали AB с каждой из двух оптических осей волновых нормалей. Так как $v_p=c/n$, $v_x=c/n_x$ и т. д., то (44) принимает вид

$$\left(\frac{1}{(n')^2}-\frac{1}{(n'')^2}\right)=\left(\frac{1}{n_x^2}-\frac{1}{n_z^2}\right)\sin\vartheta_1\sin\vartheta_2, \tag{45}$$

или приближенно, так как разности показателей преломления малы по сравнению с их величинами,

$$n''-n'=(n_z-n_x)\sin\vartheta_1\sin\vartheta_2. \tag{46}$$

Подставляя последнее соотношение в (31) и снова вводя величину $h/\cos\vartheta_2=\rho$, получим для разности фаз

$$\delta=\frac{2\pi\rho}{\lambda}(n_z-n_x)\sin\vartheta_1\sin\vartheta_2. \tag{47}$$

Как мы видим, поверхности постоянной разности фаз определяются теперь выражением

$$\rho\sin\vartheta_1\sin\vartheta_2=C \quad (C=\text{const}). \tag{48}$$

В направлении каждой оптической оси ($\vartheta_1=0$ или $\vartheta_2=0$) ρ стремится к бесконечности, так что эти поверхности асимптотически приближаются к цилиндрам, окружающим оптические оси. Когда ϑ_1 мало, ϑ_2 приблизительно равно углу 2β между двумя оптическими осями, и (48) тогда переходит в соотношение

$$\rho\sin\vartheta_1=\frac{C}{\sin 2\beta}. \tag{49}$$

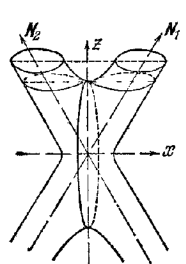


Рис. 14.25 Поверхность постоянной разности фаз для оптически двухосного кристалла

Но $\rho\sin\vartheta_1$ — это расстояние точки на поверхности до оптической оси $\vartheta_1=0$, а $\rho\sin\vartheta_2$ — расстояние до оси $\vartheta_2=0$. Следовательно, «асимптотические цилиндры» являются круговыми. Форма поверхности постоянной разности фаз в общем случае показана на рис. 14 25. Очевидно, что вблизи оптических осей изохроматы являются замкнутыми кривыми, приближающимися к эллипсам и окружающими две точки в фокальной плоскости, которые соответствуют оптическим осям.

Главные изогиры получают, как и раньше, находя направления волновых нормалей, для которых направления колебаний в кристалле совпадают с направлениями, пропускаемыми призмами Николя. Мы можем использовать построение эллипсоида волновых нормалей, выполненное в п. 14 2.3, *а*, где было показано, что плоскости колебаний делят пополам углы между плоскостями (N_1, s) и (N_2, s); здесь N_1 и N_2 — направления оптических осей нашего кристалла. Таким образом, если у кристалла ось z вертикальна и направления колебаний, пропускаемые призмами Николя, параллельны осям x и y, то главные изогиры будут лежать в плоскостях xz и yz. Следовательно, интерференционная картина будет иметь вид темного креста, причем одна пара образующих проходит через точки, соответствующие оптическим осям.

В более общем случае, когда кристалл ориентирован относительно направлений поляризаторов произвольным образом, главные изогиры проходят через точки, соответствующие оптическим осям, и имеют форму равнобочных гипербол, асимптоты которых совпадают с направлениями колебаний, пропускаемыми призмами Николя. Если при фиксированных положениях обеих призм поворачивать кристаллическую пластинку в ее плоскости, то картина изогир будет изменяться, а изохроматы (не считая их вращения) останутся такими же, так как они определяются условиями, не зависящими от направлений поляризаторов. Типичная интерференционная картина, полученная с пластинкой двухосного кристалла, показана на рис. 14.26.

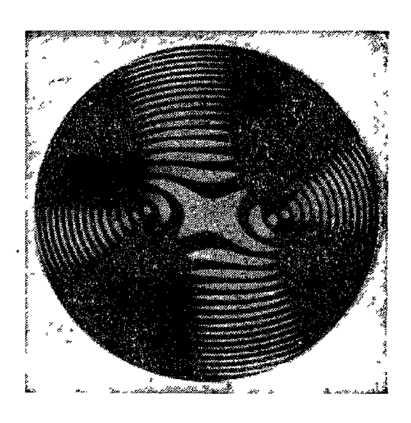


Рис. 14.26. Интерференционная картина, получающаяся с бразильским топазом.

14.4.6. Определение положения оптических осей и главных показателей преломления кристаллической среды. Так как изохроматы образуют замкнутые кривые, охватывающие оптическую ось (или оси), то наблюдение интерференционных картин сразу же позволяет установить число осей кристалла и определить их положение. Интерференционные картины можно наблюдать в микроскоп, снабженный двумя призмами Николя (так называемый *поляризационный микроскоп*), либо удаляя окуляр и фокусируя глаз на заднюю фокальную плоскость объектива (что воспроизводит условия рис. 14.21), либо помещая дополнительную линзу так, чтобы заднюю фокальную плоскость объектива можно было наблюдать через окуляр. При втором методе получается увеличенное изображение интерференционной картины и можно проводить измерения, используя калиброванную шкалу окуляра. Таким образом, можно измерять угол между оптическими осями двухосного кристалла (естественно, необходимо учитывать, что при выходе из кристалла свет преломляется). Указанные способы пригодны для определения положения оптических осей и измерения их наклона даже при наличии очень небольших кусочков кристалла, попадающихся в тонких слоях минералов.

Главные показатели преломления кристалла n_x, n_y, n_z можно определить, либо измеряя угол отклонения или полного отражения в призме, либо последовательно погружая кристалл в жидкости с разными показателями преломления.

Метод с использованием призмы более удобен для одноосных кристаллов, чем для двухосных. Призму вырезают так, чтобы ее преломляющее ребро было параллельно оптической оси волновой нормали. Тогда для обыкновенной и необыкновенной волн векторы **D** соответственно один перпендикулярен, а другой параллелен этому ребру. Оба показателя преломления можно найти по отклонениям двух лучей, выходящих из призмы, когда на одну из ее граней падает неполяризованный пучок света. Обыкновенный и необыкновенный лучи легко различить с помощью призмы Николя.

Иммерсионный метод основан на том, что прозрачное тело, погруженное в жидкость с таким же показателем преломления, становится невидимым. Поскольку при любом направлении распространения света кристалл имеет два показателя преломления, он останется видимым в любой жидкости, если его наблюдать в неполяризованном свете. Однако при использовании поляризо-

ванного света, у которого направление вектора **D** совпадает с одним из направлений колебаний в кристалле, последний станет невидимым в жидкости с соответствующим показателем преломления (n' или n''). Если направления главных диэлектрических осей известны, кристалл можно ориентировать так, чтобы свет распространялся по очереди параллельно каждой оси; при этом n' и n'' будут равны соответствующей паре главных показателей преломления n_x, n_y и n_z.

Если направления главных диэлектрических осей неизвестны, можно получить прекрасную оценку величин показателей преломления, погружая большое число кристаллов рассматриваемого типа со случайными ориентациями осей в ряд жидкостей с известными показателями преломления. Каждый кристалл станет невидимым при двух различных значениях показателя преломления и двух направлениях колебаний падающего света. Если эти показатели преломления для какого-то кристалла равны n' и n'' то, согласно неравенствам (14.3.12), получим

$$n_x \leqslant n' \leqslant n_y \leqslant n'' \leqslant n_z. \qquad (50)$$

Таким образом, n_x равно нижнему пределу значений n', n_z — верхнему пределу значений n'', а n_y равно и верхнему пределу n' и нижнему пределу n'', которые должны совпадать, если выполнено достаточное число измерений.

В случае одноосных кристаллов один из показателей преломления каждого кристалла равен n_o, а значения другого лежат в интервале между n_o и n_e.

§ 14.5. Искусственная анизотропия

14.5.1. Двойное лучепреломление, вызванное напряжением. Прозрачное изотропное вещество может стать оптически анизотропным, если его подвергнуть механическим напряжениям. Это явление, впервые обнаруженное Брюстером [25] и известное как *искусственная анизотропия, вызванная напряжением*, или *фотоупругость*, находит полезное практическое применение. Мы лишь кратко укажем, как можно использовать оптические методы для получения информации о состоянии напряжения в первоначально изотропном веществе. Предварительно мы должны получить соотношения, связывающие упругие и оптические постоянные вещества.

Напряжение и деформация в упругом твердом теле характеризуются тензорами второго ранга, тензором напряжения P_{hl} и тензором деформации r_{hl}, компоненты которых связаны друг с другом линейными соотношениями. Эти два тензора всегда симметричны, но главные оси соответствующих эллипсоидов в общем случае не совпадают с главными осями эллипсоида, соответствующего тензору диэлектрической проницаемости, который, как мы видели в § 14.1, определяет оптические свойства тела.

Когда тело подвергнуто напряжению, тензор его диэлектрической проницаемости изменяется, и в первом приближении можно предположить, что *изменения* компонент этого тензора линейно связаны с шестью компонентами тензора напряжения и, следовательно, с шестью компонентами тензора деформации. Таким образом, мы приходим к необходимости ввести две новые группы коэффициентов, связанных с напряжением и деформацией и характеризующих указанные соотношения. Эти коэффициенты называются *упругооптическими коэффициентами напряжения* и *деформации* соответственно.

Если мы свяжем оси координат с главными диэлектрическими осями ненапряженного материала, то эллипсоид волновых нормалей будет определяться уравнением

$$\frac{x^2}{\varepsilon_x} + \frac{y^2}{\varepsilon_y} + \frac{z^2}{\varepsilon_z} = 1. \qquad (1)$$

При наличии напряжения с компонентами P_{xx}, P_{xy}, ... этот эллипсоид перейдет в другой, уравнение которого можно записать в виде

$$a_{xx}x^2+a_{yy}y^2+a_{zz}z^2+a_{yz}yz+a_{zx}zx+a_{xy}xy=1. \qquad (2)$$

В соответствии с нашим предположением каждый коэффициент a_{kl} отличается от соответствующего коэффициента в (1) на линейную функцию от P. Таким образом, мы имеем шесть соотношений; приведем два типичных соотношения:

$$\left.\begin{aligned} a_{xx}-\frac{1}{\varepsilon_x}&=q_{11}P_{xx}+q_{12}P_{yy}+q_{13}P_{zz}+q_{14}P_{yz}+q_{15}P_{zx}+q_{16}P_{xy},\\ a_{yz}&=q_{41}P_{xx}+q_{42}P_{yy}+q_{43}P_{zz}+q_{44}P_{yz}+q_{45}P_{zx}+q_{46}P_{xy}.\end{aligned}\right\} \qquad (3)$$

В этой записи каждая цифра 1—6 индекса относится к паре осей; таким образом, $1=xx$, $2=yy$, $3=zz$, $4=yz$, $5=zx$, $6=xy$. Аналогичная система уравнений связывает коэффициенты a_{kl} с компонентами тензора деформации.

Влияние напряжения на оптические свойства можно также выразить через деформацию лучевого эллипсоида, и тогда получатся еще две системы линейных уравнений с 36 коэффициентами. Эти коэффициенты связаны с коэффициентами эллипсоида волновых нормалей, так как направления главных осей обоих эллипсоидов всегда совпадают, а величины полуосей одного являются обратными величинами полуосей другого.

Соотношения (3) принимают более простую форму, если изучаемая структура обладает элементами симметрии. Для кристаллов *кубической системы* все три главные оси x, y и z эквивалентны, и поэтому между упруго-оптическими коэффициентами напряжения выполняются соотношения *)

$$q_{11}=q_{22}=q_{33}, \qquad q_{12}=q_{21}=q_{23}=q_{32}=q_{31}=q_{13}, \qquad q_{44}=q_{55}=q_{66}, \qquad (4)$$

причем все остальные коэффициенты равны нулю.

Для *изотропных веществ* соотношения (3) должны оставаться неизменными при любом изменении осей. Это возможно только в том случае, если упруго-оптические коэффициенты напряжения удовлетворяют условиям (4) и, кроме того, выполняется соотношение

$$2q_{44}=q_{11}-q_{12}. \qquad (5)$$

Таким образом, в данном случае имеются лишь две независимые постоянные. Так как все системы осей теперь эквивалентны, мы можем использовать любые оси координат и, в частности, главные оси тензора напряжений; тогда $P_{yz}=P_{zy}=P_{xy}=0$, и вместо (3) мы получим более простые соотношения

$$\left.\begin{aligned} a_{xx}-\frac{1}{\varepsilon}&=q_{11}P_{xx}+q_{12}P_{yy}+q_{12}P_{zz}, \qquad a_{yy}-\frac{1}{\varepsilon}=q_{12}P_{xx}+q_{11}P_{yy}+q_{12}P_{zz},\\ a_{zz}-\frac{1}{\varepsilon}&=q_{12}P_{xx}+q_{12}P_{yy}+q_{11}P_{zz}, \qquad a_{yz}=a_{zx}=a_{xy}=0.\end{aligned}\right\} \qquad (6)$$

Таким образом, в данном случае главные оси тензора напряжений и эллипсоида волновых нормалей совпадают, как и следовало ожидать из соображений симметрии.

Хотя кристаллы кубической симметрии, например каменная соль, при отсутствии напряжений оптически изотропны, их поведение при наличии деформации отличается от поведения истинно изотропных веществ типа стекла. Влияние напряжений удобно наблюдать, рассматривая тело, помещенное между скрещенными призмами Николя (или другими поляризующими устройствами, например поляроидными пленками, см. п. 14.6.3). Предполо-

*) Соотношения, существующие между упруго-оптическими коэффициентами напряжения для каждой кристаллической системы, рассматриваются в работах [26, 27]. Они собраны в [28].

жим, что свет падает нормально на пластинку толщиной h. Эта пластинка подвергается напряжению, приложенному так, что две главные оси тензора напряжений, например x и y, а следовательно, и тензора диэлектрической проницаемости лежат в плоскости пластинки и образуют углы φ и $\varphi+\pi/2$ с направлениями поляризатора и анализатора, как и в опыте, описанном в п. 14.4.3. Сечение эллипсоида нормалей плоскостью xy имеет вид эллипса

$$a_{xx}x^2+a_{yy}y^2=1, \tag{7}$$

где a_{xx} и a_{yy} определяются из (6). Показатели преломления n' и n'' для двух волн, распространяющихся в пластинке, равны

$$n'=\frac{1}{\sqrt{a_{xx}}}, \qquad n''=\frac{1}{\sqrt{a_{yy}}}. \tag{8}$$

Следовательно,

$$\frac{1}{(n')^2}-\frac{1}{(n'')^2}=a_{xx}-a_{yy}=(q_{11}-q_{12})(P_{xx}-P_{yy}). \tag{9}$$

Практически n' и n'' слабо отличаются от n_x, так что приближенно имеем

$$n''-n'=\frac{1}{2}n_x^3(q_{11}-q_{12})(P_{xx}-P_{yy}). \tag{10}$$

Подставляя это выражение в (14.4.31), получим для разности фаз δ между двумя волнами, выходящими из пластинки, соотношение

$$\delta=\frac{\pi h}{\lambda}n_x^3(q_{11}-q_{12})(P_{xx}-P_{yy}). \tag{11}$$

Таким образом, разность фаз пропорциональна величине $P_{xx}-P_{yy}$, которая представляет собой удвоенное значение напряжения сдвига в плоскости, наклоненной под углом 45° к направлениям x и y. В данном случае остается лишь *один* независимый упруго-оптический коэффициент напряжения, а именно $q_{11}-q_{12}$.

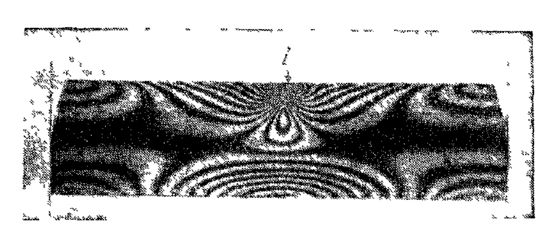


Рис. 14.27. Картина напряжений в бруске, возникающая под действием концентрированной нагрузки.

Положение груза указано стрелкой L, концы бруска (не показаны) закреплены.

Отсюда следует, что если напряженную пластинку стекла или прозрачной пластмассы поместить между скрещенными призмами Николя, будут видны светлые и темные полосы, и эти полосы соответствуют контурам равного напряжения сдвига. Такая «картина напряжений» показана на рис. 14.27. В любом данном участке картины наибольшая контрастность полос наблюдается при угле между главными осями напряженной системы в этом участке и направлениями колебаний, пропускаемых призмами Николя, равном 45°. Если главные оси тензора напряжения параллельны направлениям, пропускаемым призмами Николя, полосы исчезают и поле зрения становится темным. Таким образом, направления осей напряженной системы можно определить, поворачивая скрещенные поляризаторы, тогда как величину напряжения сдвига можно найти, зная порядок полос. Этот метод используется для исследования напряжений в промышленных конструкциях; модель конструкции изготовляют из подходящей пластмассы и описанным выше способом наблюдают непосредственно влияние напряжений. Указанный способ часто позволяет избежать очень трудоемких расчетов *).

*) Подробное описание этого метода приведено, например, в книге [29] или [30].

14.5.2. Двойное лучепреломление формы. Способность кристаллов к двойному лучепреломлению можно объяснить, исходя из анизотропии электрических свойств молекул, составляющих кристаллы. Однако двойное лучепреломление может возникать и вследствие анизотропии элементов, значительно более крупных, чем молекулы. Речь идет в данном случае о некоторой упорядоченной системе частиц из оптически изотропного вещества, размер которых велик по сравнению с размерами молекул, но мал по сравнению с длиной волны света. Тогда говорят о *двойном лучепреломлении формы*.

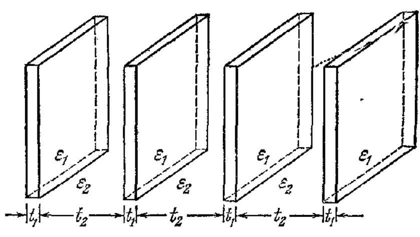


Рис 14 28. Регулярная система тонких параллельных пластин.

Часто из оптических измерений можно получить информацию о субмикроскопических частицах, которые обусловливают двойное лучепреломление указанного типа. Мы поясним принцип этого метода, рассмотрев некоторый идеализированный случай регулярной совокупности частиц, которые имеют форму тонких параллельных пластин. Пусть t_1 — толщина каждой пластины, а t_2 — ширина промежутков между ними (рис. 14.28). Далее пусть ε_1 — диэлектрическая проницаемость каждой из пластин, а ε_2 — диэлектрическая проницаемость окружающей их среды. Предположим, что на эту систему падает плоская монохроматическая волна и что ее электрический вектор перпендикулярен к пластинам Если предположить, что линейные размеры граней пластин велики, а толщины t_1 и t_2 малы по сравнению с длиной волны, то поле в пластинах и между ними можно считать однородным. Далее, в соответствии с изложенным в п. 1.1.3, нормальная составляющая вектора электрического смещения должна оставаться непрерывной при пересечении поверхности, на которой резко меняются свойства среды. Следовательно, вектор электрического смещения $\mathbf{D}$ должен быть одинаковым как внутри пластин, так и вне их. Если $\mathbf{E}_1$ и $\mathbf{E}_2$ — соответствующие электрические поля, т. е.

$$\mathbf{E}_1 = \frac{\mathbf{D}}{\varepsilon_1}, \qquad \mathbf{E}_2 = \frac{\mathbf{D}}{\varepsilon_2}, \tag{12}$$

то среднее поле $\mathbf{E}$, полученное усреднением по всему объему, равно

$$\mathbf{E} = \frac{t_1 \frac{\mathbf{D}}{\varepsilon_1} + t_2 \frac{\mathbf{D}}{\varepsilon_2}}{t_1 + t_2}. \tag{13}$$

В этом случае эффективная диэлектрическая проницаемость $\varepsilon_\perp$ равна

$$\varepsilon_\perp = \frac{D}{E} = \frac{(t_1 + t_2)\,\varepsilon_1\varepsilon_2}{t_1\varepsilon_2 + t_2\varepsilon_1} = \frac{\varepsilon_1\varepsilon_2}{f_1\varepsilon_2 + f_2\varepsilon_1}, \tag{14}$$

где $f_1 = t_1/(t_1 + t_2)$ и $f_2 = t_2/(t_1 + t_2) = 1 - f_1$ — доли общего объема, занимаемые пластинами и окружающей средой соответственно.

Предположим теперь, что электрический вектор падающего поля параллелен пластинам. Согласно п. 1.1.3 тангенциальная составляющая электрического вектора непрерывна на поверхности разрыва, так что в данном случае он будет иметь одно и то же значение $\mathbf{E}$ внутри пластин и между ними. Электрические смещения в обоих этих областях равны

$$\mathbf{D}_1 = \varepsilon_1\mathbf{E}, \qquad \mathbf{D}_2 = \varepsilon_2\mathbf{E} \tag{15}$$

и, значит, для среднего электрического смещения $\mathbf{D}$ имеем

$$\mathbf{D} = \frac{t_1\varepsilon_1\mathbf{E} + t_2\varepsilon_2\mathbf{E}}{t_1 + t_2}. \tag{16}$$

Следовательно, в этом случае эффективная диэлектрическая проницаемость окажется равной

$$\varepsilon_{\parallel}=\frac{D}{E}=\frac{t_1\varepsilon_1+t_2\varepsilon_2}{t_1+t_2}=f_1\varepsilon_1+f_2\varepsilon_2. \qquad (17)$$

Так как эффективная диэлектрическая проницаемость одинакова для всех направлений, параллельных пластинам, но различна для направлений, перпендикулярных к ним, наша система ведет себя как одноосный кристалл с оптической осью, нормальной к плоскости пластин. Разность $\varepsilon_{\parallel}-\varepsilon_{\perp}$ всегда положительна, так как, согласно (14) и (17),

$$\varepsilon_{\parallel}-\varepsilon_{\perp}=\frac{f_1f_2(\varepsilon_1-\varepsilon_2)^2}{f_1\varepsilon_2+f_2\varepsilon_1}\geqslant 0. \qquad (18)$$

Электрический вектор обыкновенной волны перпендикулярен к оптической оси, т. е. параллелен плоскости пластин. Уравнение (18) означает, что рассматриваемая система всегда ведет себя как *отрицательный одноосный кристалл* (см. п. 14.3.2). Последнее соотношение можно выразить через показатели преломления в виде

$$n_e^2-n_o^2=-\frac{f_1f_2(n_1^2-n_2^2)^2}{f_1n_2^2+f_2n_1^2}. \qquad (19)$$

Для системы не столь идеализированных частиц расчеты, естественно, более сложны *).

Значительный практический интерес представляет случай системы одинаковых тонких цилиндрических стержней, расположенных параллельно друг другу. Винер показал **), что если стержни занимают малую часть общего объема ($f_1\ll 1$), то вместо (19) мы получим

$$n_e^2-n_o^2=\frac{f_1f_2(n_1^2-n_2^2)^2}{(1+f_1)n_2^2+f_2n_1^2}. \qquad (20)$$

Иными словами, такая система ведет себя как *положительный одноосный кристалл*, причем его оптическая ось параллельна осям стержней.

Двойное лучепреломление формы используется в биологической микроскопии. Знак наблюдаемой разности указывает на близость формы частиц к форме стержня или пластинки, и если n_1 и n_2 известны, то из уравнения (19) или (20) можно оценить часть объема, занятую частицами. Для того чтобы отличить двойное лучепреломление формы от двойного лучепреломления материала самих частиц, меняют показатель преломления среды n_2. Двойное лучепреломление формы исчезнет при $n_2=n_1$, тогда как двойное лучепреломление материала не зависит от изменения n_2. При наличии двойного лучепреломления обоих типов график зависимости величины $|n_e^2-n_o^2|$ от n_2 будет иметь минимум при $n_1=n_2$, но он не будет проходить через нуль.

Упорядочение частиц, которое вызывает двойное лучепреломление формы, может быть постоянным или полупостоянным, как, например, в кристаллах вируса табачной мозаики [34]; двойное лучепреломление такого типа возникает также при временном упорядочении одинаковых частиц, взвешенных в жидкости. Подобная взвесь оптически изотропна, если частицы ориентированы случайным образом, что обычно и наблюдается в неподвижной жидкости. Если же заставить такую взвесь течь, то частицы стремятся выстроиться в определенных направлениях и наша система начнет вести себя как кристалл. Этот

*) Общую теорию см. в работе [31]. Соответствующие формулы для эллипсоидальных частиц приведены также в статье [32].

**) См. [31], стр. 581. Главные диэлектрические проницаемости прямоугольной системы параллельных цилиндров были также рассчитаны лордом Рэлеем [33]. Уравнение (20) находится в согласии с его результатами даже для значений f_1, не очень малых по сравнению с единицей, при условии, что мала разность между показателями преломления n_1 и n_2.

эффект можно наблюдать, помещая взвесь между двумя коаксиальными цилиндрами: одним вращающимся и другим — неподвижным; затем, пропуская свет в направлении, параллельном осям цилиндров, рассматривают жидкость между скрещенными призмами Николя *). Это явление, впервые обнаруженное Максвеллом [37], помогает при исследованиях потока жидкостей, текущих мимо препятствий, давая информацию о направлении и величине градиента скорости.

§ 14.6. Поглощающие кристаллы

14.6.1. Распространение света в поглощающих анизотропных средах. Оптические свойства кристаллических сред, которые мы рассматривали до сих пор, описывались тензором диэлектрической проницаемости ε_{kl}. Для описания не только анизотропной, но и поглощающей среды необходимо дополнительно ввести *тензор проводимости* σ_{kl}. Вообще говоря, направления главных осей обоих тензоров не совпадают, и поэтому теория распространения света в таких средах довольно сложна. Однако эти направления совпадают для кристаллов высокого класса симметрии (по крайней мере для кристаллов ромбической системы). Мы ограничимся рассмотрением кристаллов такого типа, так как все существенные особенности общей теории можно проиллюстрировать с их помощью. Для таких кристаллов следует лишь заменить вещественные диэлектрические проницаемости ε_x, ε_y и ε_z комплексными проницаемостями $\hat{\varepsilon}_x$, $\hat{\varepsilon}_y$ и $\hat{\varepsilon}_z$. Мы покажем, что формально сохраняются все предыдущие формулы кристаллооптики при условии, что все величины, зависящие от $\hat{\varepsilon}_x$, $\hat{\varepsilon}_y$ и $\hat{\varepsilon}_z$, считаются комплексными.

Начнем с уравнений Максвелла для проводящей среды, а именно с уравнений

$$\operatorname{rot}\mathbf{H} = \frac{1}{c}\dot{\mathbf{D}} + \frac{4\pi}{c}\mathbf{j}, \qquad \operatorname{rot}\mathbf{E} = -\frac{1}{c}\dot{\mathbf{B}}, \tag{1}$$

и рассмотрим распространение плоской затухающей волны. В комплексной записи каждый из векторов **E**, **D**, **B**, **H** и **j** пропорционален

$$\exp\left[i\omega\left(\frac{\hat{n}}{c}(\mathbf{r}\cdot\mathbf{s}) - t\right)\right].$$

По аналогии с (14.2.3) мы получаем уравнения

$$\hat{n}\mathbf{s}\times\mathbf{H} = -\mathbf{D} + \frac{4\pi}{i\omega}\mathbf{j}, \qquad \hat{n}\mathbf{s}\times\mathbf{E} = \mathbf{B}. \tag{2}$$

Полагая $\mathbf{B} = \mu\mathbf{H}$ и исключая **H** из этих уравнений, найдем

$$\mu\left(\mathbf{D} + \frac{4\pi i}{\omega}\mathbf{j}\right) = \hat{n}^2[\mathbf{E} - \mathbf{s}(\mathbf{s}\cdot\mathbf{E})]. \tag{3}$$

Если выбрать оси координат в направлении главных осей тензора диэлектрической проницаемости (которые, по нашему предположению, совпадают с направлениями главных осей тензора проводимости), то мы получим

$$\left.\begin{aligned} D_x &= \varepsilon_x E_x, & D_y &= \varepsilon_y E_y, & D_z &= \varepsilon_z E_z,\\ j_x &= \sigma_x E_x, & j_y &= \sigma_y E_y, & j_z &= \sigma_z E_z. \end{aligned}\right\} \tag{4}$$

Подставляя выражения (4) в (3) и вводя комплексные диэлектрические проницаемости

$$\hat{\varepsilon}_k = \varepsilon_k + \frac{4\pi i}{\omega}\sigma_k \qquad (k = x,\ y,\ z), \tag{5}$$

*) Определение ориентации частиц на основании таких измерений рассмотрено в работах [35, 36].

находим вместо (3) выражение

$$\mu\varepsilon_k E_k = \hat{n}^2 [E_k - s_k (\mathbf{E}\cdot\mathbf{s})]. \tag{6}$$

Формально оно идентично соотношению (14.2.18), но вещественные постоянные ε_k и n заменены комплексными постоянными $\hat{\varepsilon}_k$ и $\hat{n}$. Переписывая последнее соотношение в форме

$$E_k = \frac{\hat{n}^2 s_k (\mathbf{E}\cdot\mathbf{s})}{\hat{n}^2 - \mu\hat{\varepsilon}_k}, \tag{7}$$

мы получим, привлекая те же соображения, что и в § 14.2 (или просто используя формальную подстановку $\varepsilon_k \to \hat{\varepsilon}_k$, $n \to \hat{n}$), *уравнение Френеля*

$$\frac{s_x^2}{\frac{1}{\hat{n}^2} - \frac{1}{\mu\hat{\varepsilon}_x}} + \frac{s_y^2}{\frac{1}{\hat{n}^2} - \frac{1}{\mu\hat{\varepsilon}_y}} + \frac{s_z^2}{\frac{1}{\hat{n}^2} - \frac{1}{\mu\hat{\varepsilon}_z}} = 0. \tag{8}$$

Вводя комплексные скорости

$$\hat{v}_p = \frac{c}{\sqrt{\mu\hat{\varepsilon}}} = \frac{c}{\hat{n}}, \qquad \hat{v}_x = \frac{c}{\sqrt{\mu\hat{\varepsilon}_x}} = \frac{c}{\hat{n}_x} \tag{9}$$

и т. д., мы снова можем записать уравнение Френеля в форме (14.2.24), т. е.

$$\frac{s_x^2}{\hat{v}_p^2 - \hat{v}_x^2} + \frac{s_y^2}{\hat{v}_p^2 - \hat{v}_y^2} + \frac{s_z^2}{\hat{v}_p^2 - \hat{v}_z^2} = 0. \tag{10}$$

Полученные выше соотношения полностью аналогичны соотношениям, относящимся к непоглощающим кристаллам; однако их физические интерпретации несколько различны. Из выражений (8) или (10) мы снова получаем квадратичное уравнение относительно $\hat{n}^2(\mathbf{s})$, т. е. находим два показателя преломления и два главных колебания $\mathbf{D}'$ и $\mathbf{D}''$, соответствующих каждому заданному направлению распространения $\mathbf{s}$. Из (7) мы видим, что отношения $D_x : D_y : D_z$ комплексны, так что в общем случае главные колебания поляризованы теперь не линейно, а по эллипсу. Еще одно отличие состоит в том, что векторы электрического смещения больше не перпендикулярны к волновой нормали $\mathbf{s}$. Действительно, из первого уравнения (2) при скалярном умножении на $\mathbf{s}$ получим уравнение

$$\mathbf{s}\cdot\mathbf{D} = \frac{4\pi}{i\omega}\mathbf{s}\cdot\mathbf{j} = \frac{4\pi}{i\omega}\left(\frac{\sigma_x}{\varepsilon_x}s_x D_x + \frac{\sigma_y}{\varepsilon_y}s_y D_y + \frac{\sigma_z}{\varepsilon_z}s_z D_z\right), \tag{11}$$

и правая его часть в общем случае не равна нулю. Однако если отношения $4\pi\sigma_k/\omega\varepsilon_k$ малы по сравнению с единицей, составляющая вектора $\mathbf{D}$ в направлении $\mathbf{s}$ мала по сравнению с самим вектором $\mathbf{D}$.

Дальнейший анализ значительно упрощается, если поглощение считается малым, т. е. если металлы исключаются, а рассматриваются лишь среды, которые до известной степени прозрачны. Этим случаем мы и ограничимся. Формально слабое поглощение означает, что можно пренебречь членами второго порядка относительно показателя затухания κ по сравнению с единицей. Поэтому, используя (9), можно записать

$$\left.\begin{aligned} \hat{n} &= n(1 + i\kappa), & \hat{n}^2 &= n^2(1 + 2i\kappa), \\ \hat{v}_p &= \frac{c}{n(1+i\kappa)} = v_p(1 - i\kappa), & \hat{v}_p^2 &= v_p^2(1 - 2i\kappa). \end{aligned}\right\} \tag{12}$$

Аналогичные выражения можно написать для каждого из индексов x, y и z, например, $\hat{n}_x = n_x(1 + i\kappa_x)$, $\hat{v}_x = c/\hat{n}_x = v_x(1 - i\kappa_x)$ и т. д. Кроме того,

$$\mu\hat{\varepsilon}_k = \hat{n}_k^2 = n_k^2(1 + 2i\kappa_k) = \mu\varepsilon_k(1 + 2i\kappa_k), \tag{13}$$

где $k = x, y, z$. Сравнение с (5) дает

$$\kappa_k = \frac{2\pi}{\omega} \frac{\sigma_k}{\varepsilon_k}. \qquad (14)$$

Возвратимся к уравнению Френеля (10) и отделим его вещественную часть от мнимой. Один из членов (10) имеет вид

$$\frac{s_x^2}{\hat{v}_p^2 - \hat{v}_x^2} = \frac{s_x^2}{v_p^2 - v_x^2 - 2i(\kappa v_p^2 - \kappa_x v_x^2)} = \frac{s_x^2}{v_p^2 - v_x^2}\left(1 + 2i\frac{\kappa v_p^2 - \kappa_x v_x^2}{v_p^2 - v_x^2}\right); \qquad (15)$$

отсюда следует, что вещественная часть (10) представляет собой уравнение Френеля в старой форме (14.2.24). Для мнимой части находим

$$\kappa v_p^2 \left\{\frac{s_x^2}{(v_p^2 - v_x^2)^2} + \frac{s_y^2}{(v_p^2 - v_y^2)^2} + \frac{s_z^2}{(v_p^2 - v_z^2)^2}\right\} = \frac{\kappa_x v_x^2 s_x^2}{(v_p^2 - v_x^2)^2} + \frac{\kappa_y v_y^2 s_y^2}{(v_p^2 - v_y^2)^2} + \frac{\kappa_z v_z^2 s_z^2}{(v_p^2 - v_z^2)^2}. \qquad (16)$$

При любом заданном направлении волновой нормали $\mathbf{s}$ в общем случае уравнение Френеля дает, как и раньше, два значения фазовой скорости v_p. Часть энергии, переносимой двумя волнами, в этом случае поглощается, и уравнение (16) дает приближенные значения для двух показателей затухания.

Выражение для κ можно записать в другом виде. Используя (4), (7), (9) и (13), найдем

$$D_k = \frac{c^2}{\mu} \frac{\varepsilon_k}{\hat{\varepsilon}_k} \frac{s_k (\mathbf{E}\cdot\mathbf{s})}{\hat{v}_k^2 - \hat{v}_p^2} = -\frac{c^2}{\mu} \frac{s_k(\mathbf{E}\cdot\mathbf{s})}{v_p^2 - v_k^2}\left[1 + 2i\frac{(\kappa - \kappa_k) v_p^2}{v_p^2 - v_k^2}\right] \qquad (k = x, y, z). \qquad (17)$$

Здесь также оставлены лишь члены первого порядка относительно показателя затухания. Теперь мнимый член в (17) содержит *разность* величин κ и во многих случаях оказывается пренебрежимо малым. Это означает, что мы *пренебрегаем эллиптичностью колебаний*. В таком приближении направления двух векторов $\mathbf{D}$, принадлежащих любой данной волновой нормали $\mathbf{s}$, совпадают с их направлениями для непоглощающего кристалла, который обладает такими же (вещественными) главными диэлектрическими проницаемостями. Тогда имеем

$$\frac{s_k^2}{(v_p^2 - v_k^2)^2} = \left(\frac{\mu}{c^2}\right)^2 \frac{D_k^2}{(\mathbf{s}\cdot\mathbf{E})^2}, \qquad (18)$$

и (16) можно переписать в виде

$$\kappa v_p^2 = \frac{\kappa_x v_x^2 D_x^2 + \kappa_y v_y^2 D_y^2 + \kappa_z v_z^2 D_z^2}{\mathbf{D}^2}. \qquad (19)$$

Очевидно, что эта формула может стать неверной, если v_p приблизится к одной из главных скоростей v_k, так как мнимая часть в (17), которой мы пренебрегли, содержит в знаменателе разность $v_p^2 - v_k^2$ (см. конец п. 14.6.3).

Поскольку коэффициенты κ' и κ'', соответствующие данному направлению волновой нормали $\mathbf{s}$, в общем случае различны, две волны поглощаются по-разному. Оба коэффициента могут зависеть от частоты и меняться различным образом с частотой; поэтому, если на кристалл будет падать белый свет, то в общем случае кристалл окажется окрашенным и его цвет будет зависеть от направления колебаний в падающем свете. Это явление называется *плеохроизмом*; в случае одноосного кристалла обычно говорят о *дихроизме*; в случае двухосного кристалла — о *трихроизме* *).

Для *одноосного кристалла* ($\hat{\varepsilon}_x = \hat{\varepsilon}_y = \hat{\varepsilon}_o$, $\hat{\varepsilon}_z = \hat{\varepsilon}_e$; кроме того, $\kappa_x = \kappa_y = \kappa_o$, $\kappa_z = \kappa_e$) все соотношения принимают более простую форму. Как и в случае

*) Происхождение этих терминов связано с тем, что для одноосного кристалла имеются два характерных цвета, а для двухосного кристалла — три. Однако ряд авторов называет дихроичным веществом любую среду, коэффициент поглощения которой зависит от состояния поляризации падающего света.

непоглощающих кристаллов, уравнение Френеля разбивается на два (см. п. 14.3.2), т. е.

$$\left.\begin{aligned}(\hat{v}_p')^2 &= \hat{v}_o^2,\\ (\hat{v}_p'')^2 &= \hat{v}_o^2\cos^2\vartheta + v_e^2\sin^2\vartheta,\end{aligned}\right\} \tag{20}$$

где, как и прежде, ϑ — угол, который волновая нормаль $\mathbf{s}$ образует с оптической осью. Первое уравнение при разделении вещественной и мнимой частей дает

$$v_p' = v_o, \qquad \kappa_p' = \kappa_o, \tag{21а}$$

а из второго находим

$$(v_p'')^2 = v_o^2\cos^2\vartheta + v_e^2\sin^2\vartheta, \qquad \kappa_p''(v_p'')^2 = \kappa_o v_o^2\cos^2\vartheta + \kappa_e v_e^2\sin^2\vartheta; \tag{21б}$$

здесь мы вновь пренебрегли членами, содержащими члены второго порядка относительно показателей затухания. Мы видим, что поглощение обыкновенной волны одинаково для всех направлений распространения.

Для *двухосного кристалла* все соотношения оказываются значительно сложнее, и мы ограничимся специальными случаями, представляющими интерес. Как и в п. 14.3.3, вначале рассмотрим те направления распространения, для которых $s_x = 0$. Тогда из уравнения Френеля (10) получаются уравнения, аналогичные (14.3.6), т. е.

$$\left.\begin{aligned}(\hat{v}_p')^2 &= \hat{v}_x^2,\\ (\hat{v}_p'')^2 &= \hat{v}_z^2 s_y^2 + \hat{v}_y^2 s_z^2.\end{aligned}\right\} \tag{22}$$

Разделяя вещественную и мнимую части, мы получим две скорости и два показателя затухания, соответствующие направлению волновой нормали $\mathbf{s}(0,\ s_y,\ s_z)$, а именно

$$(v_p')^2 = v_x^2, \qquad \kappa_p' = \kappa_x, \tag{23а}$$

$$(v_p'')^2 = v_z^2 s_y^2 + v_y^2 s_z^2, \qquad \kappa_p''(v_p'')^2 = \kappa_z v_z^2 s_y^2 + \kappa_y v_y^2 s_z^2. \tag{23б}$$

Аналогичные соотношения, конечно, выполняются и для направлений распространения, перпендикулярных к осям y и z.

В общем случае не существует вещественных значений s_y и s_z, для которых оба корня $\hat{v}_p'$ и $\hat{v}_p''$ одинаковы. Можно найти направления, для которых равны вещественные фазовые скорости v_p' и v_p'', но соответствующие показатели затухания (κ' и κ'') в общем случае не одинаковы.

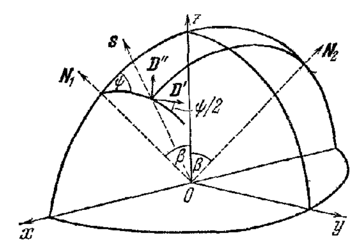


Рис. 14.29. К теории поглощающих кристаллов.

В качестве второго случая мы рассмотрим распространение света в направлениях, не очень сильно отличающихся от направления оптической оси волновых нормалей. Для того чтобы можно было применить (19), нужно определить направления $\mathbf{D}'$ и $\mathbf{D}''$. Это можно сделать, используя результат, установленный в п. 14.2.3, согласно которому две плоскости колебаний ($\mathbf{D}'$, $\mathbf{s}$) и ($\mathbf{D}''$, $\mathbf{s}$) делят пополам углы между плоскостями ($\mathbf{N}_1$, $\mathbf{s}$) и ($\mathbf{N}_2$, $\mathbf{s}$), где $\mathbf{N}_1$ и $\mathbf{N}_2$ — оси волновых нормалей. Пусть ψ — угол между плоскостью ($\mathbf{N}_1$, $\mathbf{s}$) и плоскостью xz, в которой лежат обе оптические оси. Так как плоскость ($\mathbf{N}_2$, $\mathbf{s}$) почти параллельна плоскости xz, из упомянутой выше теоремы следует, что угол между $\mathbf{D}'$ и плоскостью xz почти равен $\psi/2$ (рис. 14.29). Следовательно, проекция вектора $\mathbf{D}'$ на плоскость xz равна $D'\cos(\psi/2)$. Чтобы получить составляющую вдоль оси x, мы должны найти проекцию этого отрезка на ось x. Так как направления $\mathbf{s}$ и $\mathbf{N}_1$ приблизительно совпадают, угол между отрезком $D'\cos(\psi/2)$ и осью x почти равен углу между оптической осью $\mathbf{N}_1$ и осью z и, следователь-

но (рис. 14.30), $D'_x = D'\cos\beta\cos(\psi/2)$. Аналогичным образом определяются и другие составляющие. Итак,

$$D'_x = D'\cos\frac{\psi}{2}\cos\beta, \quad D'_y = D'\sin\frac{\psi}{2}, \quad D'_z = -D'\cos\frac{\psi}{2}\sin\beta. \tag{24а}$$

Вектор $\mathbf{D}''$ ортогонален к $\mathbf{s}$ и $\mathbf{D}'$; его составляющие сразу же можно получить, заменяя в (24а) $\psi/2$ на $\psi/2+\pi/2$, что дает

$$D''_x = -D''\sin\frac{\psi}{2}\cos\beta, \quad D''_y = D''\cos\frac{\psi}{2}, \quad D''_z = D''\sin\frac{\psi}{2}\sin\beta. \tag{24б}$$

Подставим два последних выражения в (19) и используем приближение $v'_p = v''_p = v_y$ $(v_x > v_y > v_z)$, которое вполне оправдано, поскольку мы ограничиваемся направлениями, не слишком отличающимися от направления оптической оси. Таким образом, мы получаем для искомых показателей затухания κ' и κ'' соотношения

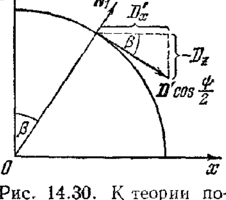


Рис. 14.30. К теории поглощающих кристаллов.

$$\left.\begin{aligned} \kappa' v_y^2 &= (\kappa_x v_x^2\cos^2\beta + \kappa_z v_z^2\sin^2\beta)\cos^2\frac{\psi}{2} + \kappa_y v_y^2\sin^2\frac{\psi}{2}, \\ \kappa'' v_y^2 &= (\kappa_x v_x^2\cos^2\beta + \kappa_z v_z^2\sin^2\beta)\sin^2\frac{\psi}{2} + \kappa_y v_y^2\cos^2\frac{\psi}{2}. \end{aligned}\right\} \tag{25}$$

Учитывая приближение, в котором получена формула (19), нельзя ожидать, что эти формулы останутся справедливыми, когда направление $\mathbf{s}$ будет мало отличаться от направления оптической оси. В предельном случае, когда волновая нормаль направлена вдоль оптической оси, угол ψ становится неопределенным. Для получения в этом случае соответствующих показателей затухания возвратимся к уравнению Френеля. При любом направлении $\mathbf{s}$ в плоскости xz $(s_y = 0)$ мы получим, как в уравнениях (23), приравнивая вещественные и мнимые части,

$$\left.\begin{aligned} (v'_p)^2 &= v_y^2, & \kappa'_p &= \kappa_y, \\ v''_p &= v_x s_z^2 + v_z s_x^2, & \kappa''_p (v''_p)^2 &= \kappa_x v_x^2 s_z^2 + \kappa_z v_z^2 s_x^2. \end{aligned}\right\} \tag{26}$$

В частности, для оптической оси $s_x = \sin\beta$, $s_z = \cos\beta$, где β — угол между каждой из оптических осей и осью z, определяемый (14.3.11). Кроме того, $v'_p = v''_p = v_y$ и правые части уравнений (26) примут вид (если мы еще напишем $\kappa_\perp$ вместо κ' и $\kappa_\parallel$ вместо κ'')

$$\kappa_\perp = \kappa_y, \quad \kappa_\parallel v_y^2 = \kappa_x v_x^2\cos^2\beta + \kappa_z v_z^2\sin^2\beta. \tag{27}$$

Здесь $\kappa_\perp$ — показатель затухания для волны $\mathbf{D}$, поляризованной в плоскости, перпендикулярной к плоскости оптических осей, а $\kappa_\parallel$ — показатель затухания для волны $\mathbf{D}$, поляризованной в плоскости осей. Таким образом, поглощение волны, распространяющейся в направлении оптической оси, зависит от направления ее колебаний.

Удобно выразить κ' и κ'' через показатели $\kappa_\parallel$ и $\kappa_\perp$ и азимут поляризации ψ, подставляя (27) в (25). Это дает

$$\left.\begin{aligned} \kappa' &= \kappa_\parallel\cos^2\frac{\psi}{2} + \kappa_\perp\sin^2\frac{\psi}{2} = \frac{\kappa_\parallel+\kappa_\perp}{2} + \frac{\kappa_\parallel-\kappa_\perp}{2}\cos\psi, \\ \kappa'' &= \kappa_\parallel\sin^2\frac{\psi}{2} + \kappa_\perp\cos^2\frac{\psi}{2} = \frac{\kappa_\parallel+\kappa_\perp}{2} - \frac{\kappa_\parallel-\kappa_\perp}{2}\cos\psi. \end{aligned}\right\} \tag{28}$$

14.6.2. Интерференционные картины, получающиеся с пластинками поглощающих кристаллов. Теперь кратко рассмотрим интерференционные эффекты, возникающие с пластинками поглощающих кристаллов, вырезанными перпендикулярно к оптической оси волновых нормалей. Теория в этом случае

не очень сильно отличается от теории, относящейся к пластинкам из непоглощающих кристаллов. Единственное важное различие заключается в том, что два интерферирующих луча поглощаются различным образом, в результате чего видность полос уменьшается. Другие заключения (и, в частности, выражение для разности фаз в нашем приближении) остаются неизменными, так как геометрические законы распространения не изменяются.

Согласно (13.1.19) амплитуда плоской волны, распространившейся в поглощающей среде на расстояние l, уменьшается в $\exp\{-\omega n\kappa l/c\}$ раз. Следовательно, для такого же случая, как в п. 14.4.3 и при тех же обозначениях (см. рис. 14.20) амплитуды главных колебаний при выходе из пластинки определяются выражениями (см. (14.4.13))

$$OB = E\exp\left(-(\omega\kappa'/v')l\right)\cos\varphi, \qquad OC = E\exp\left(-(\omega\kappa''/v'')l\right)\sin\varphi. \tag{29}$$

Здесь $l = h/\cos\theta_2$, где h — толщина пластинки и θ_2 — угол, который волновая нормаль в пластинке образует с осью волновых нормалей. Предполагается, что два пути в пластинке одинаковы для обеих волн. Это приблизительно верно, если мы ограничимся направлениями волновых нормалей, близкими к оптической оси. В том же приближении мы можем положить в (29) $v' = v'' = v_y$. Удобно положить

$$u = \omega l/v_y \approx \omega l/v' \approx \omega l/v''. \tag{30}$$

Тогда получим

$$OB = E\exp(-\kappa' u)\cos\varphi, \qquad OC = E\exp(-\kappa'' u)\sin\varphi. \tag{31}$$

Для амплитуд волн, прошедших поляризатор и анализатор, находим вместо (14.4.14) (см. рис. 14.20) следующие выражения:

$$\left.\begin{aligned} OF &= E\exp(-\kappa' u)\cos\varphi\cos(\varphi-\chi), \\ OG &= E\exp(-\kappa'' u)\sin\varphi\sin(\varphi-\chi). \end{aligned}\right\} \tag{32}$$

Полная интенсивность света, получающаяся в результате интерференции, равна

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta, \tag{33}$$

где, с точностью до несущественных коэффициентов пропорциональности, $I_1 = (OF)^2$, $I_2 = (OG)^2$, а разность фаз δ рассчитывается, как и раньше.

Ниже мы рассмотрим некоторые случаи, представляющие особый интерес.

а. Одноосные кристаллы. В этом случае мы имеем

$$v_x = v_y = v_o, \qquad v_z = v_e, \qquad \kappa_x = \kappa_y = \kappa_o, \qquad \kappa_z = \kappa_e. \tag{34}$$

Направление колебаний в необыкновенном луче лежит в главной плоскости, т. е. в плоскости, в которой лежат волновая нормаль и оптическая ось. Поэтому мы можем отождествить угол между вектором **D** для необыкновенной волны и направлением поляризатора OP с углом φ в уравнении (32). Однако, для того чтобы сохранилось согласие с (21а) и (21б), необходимо поменять местами κ' и κ''. При *скрещенных* призмах Николя ($\chi = \pi/2$) получим из (32)

$$OF = E\exp(-\kappa' u)\cos\varphi\sin\varphi, \qquad OG = -E\exp(-\kappa'' u)\sin\varphi\cos\varphi, \tag{35}$$

а (33) дает

$$I = {}^1/_4 E^2\sin^2 2\varphi\,\{\exp(-2\kappa' u) + \exp(-2\kappa'' u) - 2\exp[-(\kappa'+\kappa'')u]\cos\delta\}. \tag{36а}$$

Для оптической оси $\kappa' = \kappa''$, $\delta = 0$, и (36а) принимает вид

$$I_0 = 0. \tag{36б}$$

Следовательно, центр картины (точка, соответствующая оптической оси) оказывается темным. Поле зрения пересекается темной изогирой, определяемой соотношением $\sin 2\varphi = 0$, т. е. соотношениями $\varphi = 0$ и $\varphi = \pi/2$. Таким образом, изогира имеет форму креста, образующие которого параллельны направлениям поляризатора и анализатора (см. п. 14.4.4). Минимум и максимум

интенсивности равны

$$\left.\begin{aligned} I_{\text{мин}} &= \frac{E^2}{4}\sin 2\varphi\,\{\exp(-2\kappa' u)+\exp(-2\kappa'' u)-2\exp[-(\kappa'+\kappa'')u]\},\\ I_{\text{макс}} &= \frac{E^2}{4}\sin 2\varphi\,\{\exp(-2\kappa' u)+\exp(-2\kappa'' u)+2\exp[-(\kappa'+\kappa'')u]\},\end{aligned}\right\} \quad (37)$$

и, значит, для видности полос $\mathscr{V}$ можем написать

$$\mathscr{V} = \frac{I_{\text{макс}}-I_{\text{мин}}}{I_{\text{макс}}+I_{\text{мин}}} = \frac{2\exp[-(\kappa'+\kappa'')u]}{\exp(-2\kappa' u)-\exp(-2\kappa'' u)} = \frac{1}{\operatorname{ch}\{(\kappa'-\kappa'')u\}}. \quad (38)$$

Таким образом, видность тем больше, чем меньше разность между κ' и κ''. Так как для направлений волновых нормалей, близких к направлению оптической оси (угол ϑ мал), κ' и κ'' почти равны, то в этой области полосы должны быть ясно видны при условии, что наша пластинка пропускает достаточное количество света. Если κ_o мал по сравнению κ_e (например, для цианида магния — платины), то вблизи оптической оси будет относительно слабое поглощение и центральные полосы окажутся яркими. Если κ_o велик по сравнению с κ_e (например, для турмалина), то поглощение вблизи оптической оси будет относительно наибольшим и центральные полосы окажутся темными. В обоих случаях видность полос уменьшается от центра к краю поля. Если κ_o и κ_e почти равны, то интерференционная картина будет подобна картине для непоглощающего кристалла, но с увеличением расстояния от центра видность полос будет уменьшаться.

б. Двухосные кристаллы. Мы вновь ограничимся случаем скрещенных николей ($\chi = \pi/2$) и предположим, что грани пластинки перпендикулярны к одной из оптических осей (N_1). Мы рассмотрим лишь волны, направления распространения которых близки к этой оси.

Пусть N_1, N_2 и Q — точки пересечения двух оптических осей и волновой нормали с выходной гранью пластинки соответственно. Пусть далее ψ — угол между линией N_1Q и продолжением линии N_2N_1, а α — угол между плоскостью оптических осей и плоскостью колебаний, пропускаемых поляризатором P.

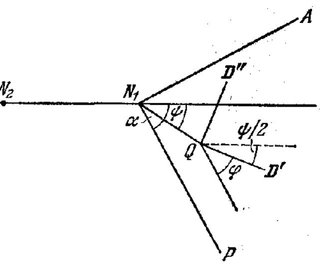


Рис. 14.31. К теории интерференции в пластинках поглощающих двухосных кристаллов.

Если $N_1Q \ll N_1N_2$, то угол, который направление колебаний $\mathbf{D}'$ образует с N_1N_2, приблизительно равен $\psi/2$, так что угол φ между $\mathbf{D}'$ и направлением N_1P приблизительно равен (рис. 14.31)

$$\varphi = \alpha - \psi/2 \quad (39)$$

Тогда для амплитуд (32) интерферирующих волн имеем

$$\left.\begin{aligned} OF &= E\cos\left(\alpha-\frac{\psi}{2}\right)\sin\left(\alpha-\frac{\psi}{2}\right)\exp(-\kappa' u),\\ OG &= -E\sin\left(\alpha-\frac{\psi}{2}\right)\cos\left(\alpha-\frac{\psi}{2}\right)\exp(-\kappa'' u).\end{aligned}\right\} \quad (40)$$

Следовательно, интенсивность запишется в виде

$$I = \frac{E^2}{4}\sin^2(2\alpha-\psi)\,\{\exp(-2\kappa' u)+\exp(-2\kappa'' u)-2\exp[-(\kappa'+\kappa'')u]\cos\delta\}. \quad (41a)$$

Для волны, распространяющейся в направлении оптической оси, угол ψ неопределен. В этом специальном случае мы разложим колебания на составляющие, параллельные и перпендикулярные к плоскости, содержащей оптические оси, т. е. положим $\psi = 0$ и напишем $\kappa_{\parallel}$ и $\kappa_{\perp}$ вместо κ' и κ'' в (41a) (см. (28)),

Поскольку, кроме того, $\delta=0$, вместо (41а) мы получим

$$I_0=\frac{E^2}{4}\sin^2 2\alpha\,\{\exp(-\kappa_\parallel u)-\exp(-\kappa_\perp u)\}^2. \qquad (41б)$$

Из уравнения (41а) следует, что интенсивность равна нулю вдоль линии, для которой $\sin(2\alpha-\psi)=0$, т. е. эта линия служит главной изогирой. В случае непоглощающего кристалла главная изогира проходила через точку, соответствующую оптической оси; теперь это не так, если только, как мы видим из (41б), не выполняется равенство $\alpha=0$ или $\alpha=\pi/2$, т. е. если плоскость поляризатора не параллельна или перпендикулярна плоскости, содержащей обе оптические оси.

Темные изохроматы, определяемыми уравнением $\cos\delta=1$, т. е. значениями $\delta=0,\ \pm2\pi,\ \pm4\pi,\ \ldots$, имеют вид колец, охватывающих точку, которая соответствует оптической оси. Видность полос, снова определяемая выражением (38), отлична от нуля лишь при почти равных κ' и κ''. Согласно (28) для этого нужно, чтобы угол ψ мало отличался от $\pi/2$ или $-\pi/2$.

Посмотрим теперь, как меняется интенсивность в зависимости от ψ при фиксированном δ. Если воспользоваться соотношением (28), то уравнение (41а) можно переписать в виде

$$I=\frac{E^2}{2}\exp\left[-(\kappa_\parallel+\kappa_\perp)u\right]\sin^2(2\alpha-\psi)\,\{\mathrm{ch}\left[(\kappa_\parallel-\kappa_\perp)u\cos\psi\right]-\cos\delta\}. \qquad (42)$$

Член в фигурных скобках достигает наибольшего значения при максимальной величине $|\cos\psi|$ и наименьшего при минимальной величине $|\cos\psi|$. Следовательно, интенсивность максимальна, когда $\psi=0$ и $\psi=\pi$, и минимальна, когда $\psi=\pi/2$ и $-\pi/2$. Таким образом, поле зрения пересекается, помимо темной изогиры $\psi=2\alpha$, еще и темными полосами, перпендикулярными к плоскости, содержащей оптические оси *).

14.6.3. Дихроичные поляризаторы. В предыдущем разделе мы изучали распространение поляризованного света через пластинку поглощающего кристалла. Здесь мы рассмотрим ее влияние на естественный свет.

Мы снова предположим, что грани пластинки перпендикулярны оптической оси волновых нормалей (или одной такой оси, если кристалл двухосный), и будем рассматривать распространение света в направлении, близком к направлению оптической оси. Параллельный пучок естественного света можно считать состоящим из двух взаимно некогерентных пучков с равными амплитудами, поляризованных в любых двух взаимно ортогональных направлениях, перпендикулярных к направлению распространения (см. (10.8.56)). В качестве направлений колебаний в этих двух составляющих пучках выберем направления главных колебаний в кристалле. Если E — амплитуда колебаний в каждом из таких пучков на входе пластинки, то их амплитуды после прохождения в пластинке расстояния l будут равны

$$E'=E\exp(-\kappa' u),\qquad E''=E\exp(-\kappa'' u), \qquad (43)$$

где, как и раньше, $u=\omega l/v_y$, причем считается, что свет квазимонохроматичен и средняя частота его ω. Тогда для полной интенсивности получим

$$I=I'+I'', \qquad (44)$$

где

$$I'=I_0\exp(-2\kappa' u),\qquad I''=I_0\exp(-2\kappa'' u), \qquad (45)$$

и $I_0=E^2$. Для направления, совпадающего с оптической осью, κ' и κ'' необходимо заменить на $\kappa_\parallel$ и $\kappa_\perp$.

*) Из уравнения (44), которое с помощью (28) можно переписать в виде

$$I=2E^2\exp\left[-(\kappa_\parallel+\kappa_\perp)u\right]\mathrm{ch}\,\{(\kappa_\parallel-\kappa_\perp)u\cos\psi\},$$

следует, что эти темные полосы появляются также в отсутствие поляризатора и анализатора при прохождении через пластинку естественного света.

Мы видим, что после прохождения света в среде на расстояние l отношение амплитуд составляющих равно $\exp\{-(\kappa'-\kappa'')u\}$ и, следовательно, свет станет частично поляризованным. Если два показателя поглощения такого вещества различаются очень сильно, то относительно тонкой пластинки его уже достаточно для преобразования

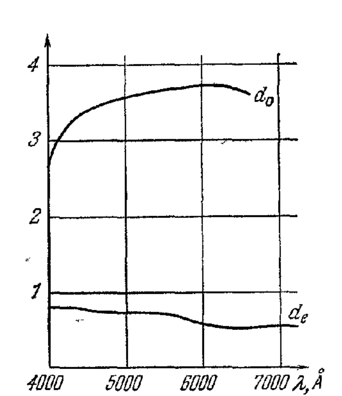


Рис. 14.32. Дихроизм пластинки из черного турмалина толщиной около 0,2 *мм*, вырезанной под углом 24° к оптической оси, для света, падающего перпендикулярно к оптической оси [38].

$d=4\pi\,(\lg e)\,(nkl/\lambda)\approx 5{,}5nkl/\lambda$

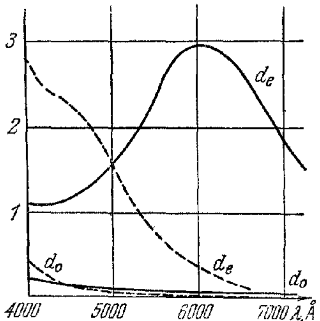


Рис. 14.33. Дихроизм иода в ориентированном поливиниловом спирте [38].

Голубой образец — сплошная кривая, коричневый — пунктирная.

$d=4\pi\,(\lg e)\,(nkl/\lambda)\approx 5{,}5\ nkl/\lambda$.

падающего неполяризованного пучка в почти линейно поляризованный пучок, т. е. такая пластинка действует как *поляризатор*. Примером природного кристалла, являющегося поляризатором такого типа, служит турмалин, в котором обыкновенный луч поглощается значительно сильнее необыкновенного. Однако для большинства длин волн турмалин сильно поглощает и необыкновенный луч (рис. 14.32), так что этот кристалл не очень подходит для практических применений. В результате исследований главным образом Лэнда и его сотрудников (приблизительно 1932 г.) появилась возможность изготовлять синтетические дихроичные материалы, служащие прекрасными поляризаторами. Эти материалы, известные в обиходе как *поляроиды*, не являются монокристаллами, а представляют собой пластинки органических полимеров с молекулами в виде длинных цепочек, почти полностью выстроившихся параллельно друг другу в результате растяжения или какой-либо другой обработки *). Такие пластинки иногда имеют окраску. Полимером, наиболее подходящим для этой цели, является синтетический поливиниловый спирт $(-CH_2-CHOH-)_x$. На рис. 14.33 показана зависимость дихроизма иода в ориентированном поливиниловом спирте от длины волны.

Отношение коэффициентов поглощения хорошего дихроичного поляризатора может достигать 100 : 1. Он может пропускать около 80% света, поляризованного в одном направлении, и менее 1% света, поляризованного в направлении, перпендикулярном первому. Можно получать большие листы таких поляризующих материалов; существуют промышленные машины, вырабатывающие непрерывную ленту шириной около 75 *см*. В этом отношении такие поляроидные пленки превосходят призму Николя, так как большие куски исландского шпата хорошего оптического качества встречаются очень редко.

Наконец, напомним, что в основной части настоящей главы мы предполагали, что можно пренебречь эллиптичностью главных световых колебаний

*) Более полное обсуждение дихроичных поляризаторов проводится в [38] или в [39].

в кристалле. Это предположение, конечно, не всегда справедливо. Можно показать, например, что в каждом поглощающем кристалле существуют четыре направления, расположенных попарно вблизи оптических осей, в которых поляризация оказывается круговой. Однако чем слабее поглощение, тем меньше области заметной эллиптичности вокруг этих специальных направлений, а сами они стремятся совпасть с оптической осью. Таким образом, и в данном случае сохраняются основные характеристики наблюдаемого явления, и поэтому нет необходимости проводить более детальное рассмотрение. Конечно, наш анализ не затрагивает некоторых тонких аспектов оптической теории поглощающих кристаллов. Для их изучения необходимо обратиться к специальной литературе (см., например, [7], стр. 861—904).

ЛИТЕРАТУРА

1. Th. Liebisch, H. Rubens, Sitzber. preuss. Akad. Wiss. 198, 876 (1919).
2. H. Rubens, Sitzber. preuss. Akad. Wiss., Phys-math. Kl., 198, 976 (1919).
3. M. Born, Optik, Springer, Berlin, 1933. (М. Борн, Оптика, ГНТИУ, 1937.)
4. R. Courant, Differential and Integral Calculus, Blackie a. Son, 1942, Vol. II. (Р. Курант, Курс дифференциального и интегрального исчисления, ГТТИ, 1933, ч. II.)
5. C. W. Bunn, Chemical Crystallography, Clarendon Press, Oxford, 1945, Ch. 2.
6. E. T. Whittaker, History of the Theories of Aether and Electricity, Vol. I: The Classical Theories, T. Nelson, London, 1951.
7. G. Szivessy, в сб. «Handbuch d. Physik», Springer, Berlin, 1928, vol. 20, p. 715.
8. M. G. Lamé, Leçons sur la théorie mathématique de l'élasticité des corps solides, Gauthier-Villars, Paris, 2me ed., 1866.
9. V. Volterra, Acta Math. 16, 153 (1892).
10. D. M Y. Sommerville, Analytical Conics, Bell a. Sons, London, 1941, p. 92.
11. G. Salmon, Analytical Geometry of Three Dimensions, Rev. by. R. A. P. Rogers, Longmans, Green a. Co., London, 1915, 5th ed. Vol. II, ch. IV.
12. J. C. Poggendorff, Pogg. Ann. 48, 461 (1839).
13. W. Haidinger, Wiener Ber. 16, 129 (1854); Pogg. Ann. 96, 486 (1855).
14. W. Voigt, Phys. Z. 6, 672, 818 (1905).
15. W. Nicol, Edinburgh New Philos. J. 6, 83 (1829).
16. L. C. Martin, An Introduction to Applied Optics, Pitman, London, 1930, Vol. 1, p. 204.
17. R W. Wood, Physical Optics, Macmillan, N. Y., 3rd ed., 1934. (Р. Вуд, Физическая оптика, ОНТИ, 1936).
18. H. G. Jerrard, J. Opt. Soc. Amer. 38, 35 (1948).
19. M. Richartz, Hsien-Yü Hsü, J. Opt. Soc. Amer. 39, 136 (1949).
20. G. Airy, Trans. Cambr. Phil. Soc. 4, 313 (1833).
21. J. Babinet, C. R. Acad. Sci. 29, 514 (1849); J. Jamin, Ann. Chim. 29, 274 (1850).
22. H. Soleil, C. R. Acad. Sci. 21, 426 (1845); 24, 973 (1847); 26, 162 (1848); J. Duboscq, H. Soleil, C. R. Acad. Sci. 31, 248 (1850).
23. M. Berek, Zbl. Miner. Geol Paläont. 388, 427, 464, 580 (1913).
24. F. Savart, Ann. d. Physik 49, 292 (1840).
25. D. Brewster, Phil. Trans. 60 (1815); 156 (1816); Trans. Roy. Soc., Edinb. 8, 369(1818).
26. F. Pockels, Ann. d. Physik 37, 158 (1889).
27. F. Pockels, Lehrbuch der Kristalloptik, Leipzig, 1906, p. 469—74.
28. G. Szivessy, в сб. «Handbuch d. Physik», Springer, Berlin, 1929, Vol. 21, p. 840.
29. E. G. Coker, L. N. Filon, A Treatise on Photoelasticity, Cambr. Univ. Press, 1931.
30. M. M. Frocht, Photoelasticity, Vol. I, 1941, Vol. II, 1948, John Wiley, N. Y. (М. М. Фрохт, Фотоупругость, Гостехиздат, т. 1, 1948, т. 2, 1950.)
31. O. Wiener, Abh. Sächs. Ges. Akad. Wiss, Math-Phys. Kl., № 6, 32, 575 (1912).
32. W. L. Bragg, A. B Pippard, Acta Cryst. 6, 865 (1953).
33 Rayleigh, Phil. Mag. 34, 481 (1892).
34. M. F Wilkins, A. Stokes, W. Seeds, G. Oster, Nature 166, 127 (1950).
35. P. Boeder, Z. f. Phys. 75, 258 (1932).
36. J. T. Edsall, в сб. «Advances in Colloidal Science», ed. E. O. Kraemer, Intersci. Publ. Inc., N. Y., 1942, Vol. I, p. 269.
37. J. C. Maxwell, Proc. Roy. Soc. 22, 46 (1873); Collected Papers, Cambr. Univ. Press, 1890, Vol. 2, p. 379.
38. E. H. Land, C. D. West, в сб. «Colloid Chemistry», ed. J. Alexander, Reinhold Publ. Corp., N. Y., 1946, Vol. 6, p. 167.
39. E. Land, J. Opt. Soc. Amer. 41, 957 (1951).
40*. В. М. Агранович, В. Л. Гинзбург, Кристаллооптика с учетом пространственной дисперсии и теория экситонов, «Наука», 1965.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ 1

ЭЛЕМЕНТЫ ВАРИАЦИОННОГО ИСЧИСЛЕНИЯ

Общая черта уравнений классической физики состоит в том, что все они могут быть выведены из вариационных принципов. Принцип Ферма в оптике (1657 г.) и принцип Мопертюи в механике (1744 г.) служат примерами наиболее ранних вариационных принципов. Из соответствующих вариационных принципов можно также получить уравнения упругости, гидродинамики и электродинамики.

Однако при рассмотрении уравнений полей, содержащих, как правило, четыре или большее число независимых переменных x, y, z, t ..., практически невозможно воспользоваться тем, что решение является стационарным значением некоторых интегралов, так как само решение дифференциальных уравнений в частных производных представляет большие трудности. В этих случаях использование вариационного принципа дает преимущество лишь при выводе законов сохранения, в частности закона сохранения энергии. Другое дело, если решаются задачи с одной независимой переменной (время в механике или длина луча в геометрической оптике). В этом случае имеют дело с обыкновенными дифференциальными уравнениями, и оказывается, что применение вариационного принципа существенно упрощает исследование решения задачи. Фактически такой подход является непосредственным обобщением обычной геометрической оптики. В своем современном виде он разработан главным образом Д. Гильбертом, и рассуждения, изложенные выше, базируются на материалах его неопубликованных лекций, прочитанных в Геттингене примерно в 1903 г. Здесь приводится теория лишь для трехмерного пространства (x, y, z), однако ее легко обобщить на многомерный случай.

1. Уравнения Эйлера как необходимое условие экстремума. Пусть $F(u, v, x, y, z)$ — заданная функция с непрерывными частными производными до второго порядка включительно по всем пяти переменным. Пусть далее C — произвольная кривая $x=x(z)$, $y=y(z)$ в пространстве x, y, z. Будем считать, что производные от x и y также непрерывны до второго порядка включительно. Если положить

$$u=x', \qquad v=y'$$

(штрих обозначает дифференцирование по z), то интеграл

$$I=\int_{z_1}^{z_2} F(x', y', x, y, z)\,dz \tag{1}$$

будет функцией $x(z)$, $y(z)$; другими словами, интеграл (1) представляет собой *функционал*. Основная задача вариационного исчисления состоит в следующем.

Найти такой участок кривой C между двумя заданными точками $P_1[x_1=x(z_1),\ y_1=y(z_1),\ z_1]$ и $P_2[x_2=x(z_2),\ y_2=y(z_2),\ z_2]$, чтобы интеграл (1) *имел экстремальное значение (минимальное или максимальное).*

Необходимые условия, которым должна удовлетворять такая кривая C, называемая *экстремалью*, выводятся с помощью простого линейного варьирования. Выберем для этого функцию $\xi(z)$ с непрерывной первой производной, которая обращается в нуль в конечных точках, т. е.

$$\xi(z_1)=\xi(z_2)=0, \tag{2}$$

и проведем новую кривую C', заменив координату x экстремали величиной $x+\varepsilon\xi$, где ε — малый параметр. Тогда уравнение (1) станет функцией ε, т. е.

$$I(\varepsilon)=\int_{z_1}^{z_2} F(x'+\varepsilon\xi', y', x+\varepsilon\xi, y, z)\,dz. \tag{3}$$

Величина

$$(\delta I)_x=\left(\frac{\partial I}{\partial\varepsilon}\right)_{\varepsilon=0}=\int_{z_1}^{z_2}\left(\frac{\partial F}{\partial x'}\xi'+\frac{\partial F}{\partial x}\xi\right)dz \tag{4}$$

называется *первой вариацией* по x. Обращение в нуль первой вариации, т. е. справедливость соотношения

$$(\delta I)_x=0 \tag{5}$$

является, очевидно, необходимым условием экстремума.

Проинтегрировав первый член в (4) по частям и воспользовавшись граничными условиями (2), получим

$$(\delta I)_x=\int_{z_1}^{z_2}\left(F_x-\frac{d}{dz}F_{x'}\right)\xi\,dz, \tag{6а}$$

где F_x обозначает $\partial F/\partial x$ и т. д. Заменяя теперь y на $y+\varepsilon\eta$, аналогичным образом получим

$$(\delta I)_y=\int_{z_1}^{z_2}\left(F_y-\frac{d}{dz}F_{y'}\right)\eta\,dz. \tag{6б}$$

Поскольку ξ и η можно выбирать произвольно в интервале $z_1\leqslant z\leqslant z_2$, то из (6а) и (6б) вытекает, что условия $(\delta I)_x=0$ и $(\delta I)_y=0$ эквивалентны следующим двум дифференциальным уравнениям, называемым *уравнениями Эйлера*

$$F_x-\frac{d}{dz}F_{x'}=0, \tag{7а}$$

$$F_y-\frac{d}{dz}F_{y'}=0. \tag{7б}$$

Соотношения (7а) и (7б) представляют собой дифференциальные уравнения второго порядка относительно $x(z)$ и $y(z)$. Оставим члены, содержащие производные наивысшего порядка, а именно:

$$F_{x'x'}x''+F_{x'y'}y''+\ldots=0, \tag{8а}$$

$$F_{y'x'}x''+F_{y'y'}y''+\ldots=0. \tag{8б}$$

Эти уравнения можно разрешить относительно x'' и y'', если соответствующий определитель не равен нулю, т. е.

$$F_{uu}F_{vv}-F_{uv}^2\neq 0. \tag{9}$$

Предположим, что это условие выполняется во всей рассматриваемой нами пятимерной области.

Решение двух дифференциальных уравнений второго порядка содержит четыре произвольные постоянные; следовательно, экстремали образуют четырехпараметрическое семейство кривых (∞^4 экстремалей).

2. Интеграл Гильберта и уравнение Гамильтона — Якоби. Чтобы исследовать свойства экстремалей, удобно рассмотреть сначала другую, но связанную с этим исследованием задачу. Будем считать, что переменные u и v зависят от x, y, z, т. е.

$$u=u(x, y, z), \qquad v=v(x, y, z). \tag{10}$$

Тогда $F[u(x, y, z), v(x, y, z), x, y, z]$ и ее частные производные F_u, F_v зависят только от x, y, z. Далее выберем кривую $C[x=x(z), y=y(z)]$ и образуем

интеграл

$$S = \int_{z_1}^{z_2} \{F + (x' - u) F_u + (y' - v) F_v\} dz. \qquad (11)$$

Задача состоит в следующем: *найти такие функции u, v, для которых интеграл S не зависит от формы кривой C*, а определяется лишь положением концевых точек P_1 и P_2, где P_1 имеет координаты $x_1 = x(z_1)$, $y_1 = y(z_1)$ и z_1; P_2 — координаты $x_2 = x(z_2)$, $y_2 = y(z_2)$, z_2. Величина S называется *интегралом Гильберта*.

Для определения u и v перепишем (11) в виде

$$S = \int_{P_1}^{P_2} (U\,dx + V\,dy + W\,dz), \qquad (12)$$

где

$$U = F_u, \qquad V = F_v, \qquad W = F - uF_u - vF_v. \qquad (13)$$

Хорошо известно, что необходимым и достаточным условием того, чтобы (12) не зависел от формы кривой интегрирования C, служит обращение в нуль компонент ротора от вектора **A**, с проекциями U, V, W, т. е.

$$\frac{\partial W}{\partial y} - \frac{\partial V}{\partial z} = 0, \qquad \frac{\partial U}{\partial z} - \frac{\partial W}{\partial x} = 0, \qquad \frac{\partial V}{\partial x} - \frac{\partial U}{\partial y} = 0. \qquad (14)$$

Мы получили систему из трех дифференциальных уравнений в частных производных относительно $u(x, y, z)$ и $v(x, y, z)$; эти уравнения, однако, не вполне независимы, поскольку для любых U, V, W справедливо тождество ($\operatorname{div} \operatorname{rot} \mathbf{A} = 0$ для любого вектора **A**)

$$\frac{\partial}{\partial x}\left(\frac{\partial W}{\partial y} - \frac{\partial V}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial U}{\partial z} - \frac{\partial W}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial V}{\partial x} - \frac{\partial U}{\partial y}\right) = 0. \qquad (15)$$

Если уравнения (14) удовлетворяются, то $U\,dx + V\,dy + W\,dz$ является полным дифференциалом, т. е.

$$dS = U\,dx + V\,dy + W\,dz, \qquad (16)$$

и S зависит только от $P_1(x_1, y_1, z_1)$ и $P_2(x_2, y_2, z_2)$. Написав для простоты x, y, z вместо x_2, y_2, z_2, получим

$$U = \frac{\partial S}{\partial x}, \qquad V = \frac{\partial S}{\partial y}, \qquad W = \frac{\partial S}{\partial z}. \qquad (17)$$

Возьмем теперь произвольную поверхность $T(x, y, z) = 0$ и в каждой ее точке P_1 проведем нормальный к ней вектор (U, V, W). Тогда, согласно (16), $dS = 0$, и, следовательно, величина

$$S(x, y, z) = S_1 \qquad (18)$$

постоянна на этой поверхности. Из двух (совместных) уравнений (13) можно определить u и v как функции координат поверхности; имеем

$$u = u(U, V, x, y, z), \qquad v = v(U, V, x, y, z). \qquad (19)$$

Если теперь решить дифференциальные уравнения (14) с такими граничными условиями, то мы получим частное решение рассматриваемой задачи, а именно решение, принимающее постоянное значение S_1 на выбранной поверхности $T(x, y, z) = 0$. Его можно получить из одного дифференциального уравнения в частных производных относительно функции S. Подставляя (19) в оставшееся уравнение (13), находим

$$W = W(U, V, x, y, z), \qquad (20)$$

или, используя (17),

$$\frac{\partial S}{\partial z} = W\left(\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}, x, y, z\right). \qquad (21)$$

Это уравнение называется *уравнением Гамильтона — Якоби* данной задачи.

Функция $S(x, y, z)$, принимающая постоянное значение на поверхности $T(x, y, z)=0$ и удовлетворяющая уравнению (21), служит решением рассматриваемой задачи. Функции u, v, обусловливающие независимость интеграла от пути интегрирования, находят из решения любых двух из (совместных) уравнений

$$F_u=\frac{\partial S}{\partial x}, \quad F_v=\frac{\partial S}{\partial y}, \qquad F-uF_u-vF_v=\frac{\partial S}{\partial z}, \tag{22}$$

полученных комбинированием (13) и (17).

3. Поле экстремалей. Установим теперь *связь между двумя задачами*, рассмотренными в пп. 1 и 2. Она состоит в следующем.

Если $u(x, y, z)$ и $v(x, y, z)$ — функции, делающие интеграл Гильберта S, определенный в (11), *не зависящим от пути, то решением дифференциальных уравнений*

$$x'=u(x, y, z), \qquad y'=v(x, y, z) \tag{23}$$

служит двухпараметрическое семейство (∞^2) экстремалей, *«ортогональных» поверхности $S(x, y, z)=S_1$.* Под «ортогональностью» мы здесь подразумеваем выполнение условия

$$U\,dx+V\,dy+W\,dz=0, \tag{24}$$

из которого следует, что вектор (U, V, W), определенный в (13) через u и v, перпендикулярен к любому элементу dx, dy, dz поверхности.

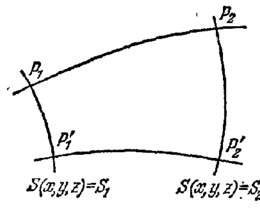


Рис. 1. К обобщению понятий волновых фронтов и лучей в геометрической оптике.

Вариационный интеграл (1) равен постоянному значению для всех экстремалей P_1P_2, $P_1'P_2'$, ..., ортогональных к поверхностям S_1 и S_2.

Рассмотрим область в пространстве x, y, z и поставим в соответствие каждой точке области вектор (u, v), непрерывный и имеющий непрерывные частные производные первого порядка. Система таких векторов, определенных в заданной области пространства, называется *полем*. Здесь мы будем говорить о поле экстремалей, величины же $u(x, y, z)$ и $v(x, y, z)$ назовем *функциями наклона* поля.

Справедлива также теорема, обратная приведенной выше. *Если поле ∞^2 экстремалей ортогонально поверхности $T(x, y, z)=0$ и u, v — его функции наклона, определенные в* (23), *то интеграл Гильберта S* (11) *не зависит от пути интегрирования.*

Прежде чем доказывать эти теоремы, отметим следующее следствие из них.

Пусть кривая C в (11) является одной из экстремалей поля; тогда интеграл Гильберта (11) сводится к вариационному интегралу $I=\int_{z_1}^{z_2} F\,dz$. Следовательно, значения этого интеграла, взятого между парами «соответствующих» точек на поверхностях $S(x, y, z)=S_1$ и $S(x, y, z)=S_2$ (т. е. между точками, расположенными на одной экстремали, ортогональной S_1 и S_2), одинаковы для всех таких пар точек (рис. 1). Поверхности $S(x, y, z)=\mathrm{const}$, и семейство ∞^2 экстремалей можно рассматривать как обобщения волновых фронтов и лучей в геометрической оптике.

Для доказательства первой теоремы рассмотрим фиксированную кривую C, которая удовлетворяет соотношению (23) и ортогональна поверхности $S(x, y, z)=S_1$, а затем применим к ней линейную вариацию, т. е. заменим x на $x+a\xi$ и y на $y+b\eta$, где a и b — малые параметры, а ξ и η — произвольные, но фиксированные дифференцируемые функции z, обращающиеся в нуль при

$z=z_1$ и $z=z_2$. По условию теоремы S не зависит от пути интегрирования, т. е.

$$\left(\frac{\partial S}{\partial a}\right)_0=0, \qquad \left(\frac{\partial S}{\partial b}\right)_0=0, \tag{25}$$

где индекс 0 обозначает, что после дифференцирования мы положили $a=b=0$.

Дифференцирование $F[u(x, y, z), v(x, y, z), x, y, z]$ дает

$$\frac{\partial F}{\partial a}=(F_u u_x+F_v v_x+F_x)\,\xi. \tag{26}$$

Из (11) с помощью (26) получим

$$\left(\frac{\partial S}{\partial a}\right)_0=\int_{z_1}^{z_2}\Big\{(F_u u_x+F_v v_x+F_x)\,\xi+ \\ +(\xi'-u_x\xi)\,F_u-v_x\xi F_v+(x'-u)\frac{\partial F_u}{\partial a}+(y'-v)\frac{\partial F_v}{\partial a}\Big\}_0 dz. \tag{27}$$

Члены, содержащие $(x'-u)$ и $(y'-v)$, обращаются в нуль, поскольку предполагается, что кривая удовлетворяет соотношению (23). Кроме того, часть членов сокращается, и мы находим

$$\left(\frac{\partial S}{\partial a}\right)_0=\int_{z_1}^{z_2}(F_x\xi+F_u\xi')\,dz, \tag{28}$$

или, интегрируя по частям,

$$\left(\frac{\partial S}{\partial a}\right)_0=\int_{z_1}^{z_2}\left(F_x-\frac{d}{dz}F_u\right)\xi\,dz; \tag{28a}$$

аналогично можем написать

$$\left(\frac{\partial S}{\partial b}\right)_0=\int_{z_1}^{z_2}\left(F_y-\frac{d}{dz}F_v\right)\eta\,dz. \tag{28б}$$

Правые части уравнений (28а) и (28б) являются первыми вариациями I (см. (6а) и (6б)), и согласно (25) они обращаются в нуль. Следовательно, кривая C удовлетворяет уравнениям Эйлера, т. е. она является экстремалью, и наша теорема доказана.

Для доказательства обратной теоремы построим поле f_1 ∞^2 экстремалей, ортогональных данной поверхности $T(x, y, z)=0$ и сравним его с другим полем f_2 ∞^2 экстремалей. Последнее поле строится следующим образом. Решаем уравнение Гамильтона — Якоби (21) с граничным условием, состоящим в том, что $S(x, y, z)$ должна быть постоянной величиной на поверхности $T(x, y, z)=0$. Если u и v определены соотношениями (22), то решение выражается интегралом (11). Тогда в соответствии с только что доказанной теоремой уравнения (23) определяют поле экстремалей f_2, ортогональных поверхности $T=0$. Однако поля f_1 и f_2 должны совпадать, поскольку они удовлетворяют одним и тем же дифференциальным уравнениям и одним и тем же граничным условиям на поверхности $T=0$. Следовательно, для данного поля f_1 величина интеграла S не зависит от пути интегрирования.

4. Нахождение всех экстремалей из решения уравнения Гамильтона — Якоби. До сих пор мы рассматривали только однопараметрическое (∞^1) семейство решений $S(x, y, z)=\text{const}$ уравнения Гамильтона — Якоби, которое соответствует двухпараметрическому (∞^2) семейству ортогональных экстремалей. Для получения полного четырехпараметрического (∞^4) семейства экстремалей необходимо рассматривать четырехпараметрическое (∞^4) семейство решений S; последнее можно получить, если поворачивать поверхность $T=0$ вокруг какой-нибудь точки и полагать в каждом случае $S(x, y, z)=S_1$. Предположим, что мы нашли «полное» решение $S(x, y, z, \alpha, \beta)$ уравнения (21), содержащее два параметра α и β. Функцию S, соответствующую любой паре

значений α и β, можно представить в виде интеграла (11), не зависящего от пути интегрирования при правильном выборе функций u и v, а именно при $u=u(x, y, z, \alpha, \beta)$ и $v=v(x, y, z, \alpha, \beta)$. Следовательно, не только S, но и $\partial S/\partial\alpha$ и $\partial S/\partial\beta$ не зависят от пути интегрирования. Поскольку $F=F[u(x, y, z, \alpha, \beta), v(x, y, z, \alpha, \beta), x, y, z]$, мы вправе написать

$$\frac{\partial F}{\partial\alpha}=F_u u_\alpha+F_v v_\alpha, \tag{29}$$

и из (11) получим

$$\left.\begin{aligned} S_\alpha&=\int_{z_1}^{z_2}\left\{(x'-u)\frac{\partial F_u}{\partial\alpha}+(y'-v)\frac{\partial F_v}{\partial\alpha}\right\}dz,\\ S_\beta&=\int_{z_1}^{z_2}\left\{(x'-u)\frac{\partial F_u}{\partial\beta}+(y'-v)\frac{\partial F_v}{\partial\beta}\right\}dz,\end{aligned}\right\} \tag{30}$$

где

$$\frac{\partial F_u}{\partial\alpha}=F_{uu}u_\alpha+F_{uv}v_\alpha, \qquad \frac{\partial F_v}{\partial\alpha}=F_{vu}u_\alpha+F_{vv}v_\alpha. \tag{31}$$

Аналогичные выражения справедливы и для других двух производных.

Так как величины интегралов (30) не зависят от пути интегрирования, выражения

$$dS_\alpha=(x'-u)\frac{\partial F_u}{\partial\alpha}+(y'-v)\frac{\partial F_v}{\partial\alpha}, \qquad dS_\beta=(x'-u)\frac{\partial F_u}{\partial\beta}+(y'-v)\frac{\partial F_v}{\partial\beta} \tag{32}$$

являются полными дифференциалами; таким образом, величины S_α и S_β — функции только концевых точек P_1 и P_2, т. е. существуют поверхности, скажем

$$\frac{\partial S(x, y, z, \alpha, \beta)}{\partial\alpha}=A, \qquad \frac{\partial S(x, y, z, \alpha, \beta)}{\partial\beta}=B, \tag{33}$$

определяемые постоянными значениями S_α и S_β; здесь A и B — постоянные. Следовательно, соотношения (33) должны быть решениями дифференциальных уравнений

$$x'=u(x, y, z, \alpha, \beta), \qquad y'=v(x, y, z, \alpha, \beta). \tag{34}$$

Отсюда следует, что на поверхностях (33) $dS_\alpha=0$, $dS_\beta=0$, и тогда из (32) вытекает (34), если соответствующий определитель не равен нулю. Согласно (31) этот определитель можно записать в виде

$$\begin{vmatrix}\dfrac{\partial F_u}{\partial\alpha} & \dfrac{\partial F_v}{\partial\alpha}\\ \dfrac{\partial F_u}{\partial\beta} & \dfrac{\partial F_v}{\partial\beta}\end{vmatrix}=\begin{vmatrix}F_{uu} & F_{uv}\\ F_{vu} & F_{vv}\end{vmatrix}\begin{vmatrix}u_\alpha & u_\beta\\ v_\alpha & v_\beta\end{vmatrix}. \tag{35}$$

Первый множитель в правой части равен выражению (9), которое, по предположению, отлично от нуля. Второй множитель обращается в нуль только в тех точках, где уравнения (34) не имеют решения для α и β, другими словами,— в точках, где экстремали пересекаются. Если исключить такие случаи, то выражения (34) будут представлять собой двухпараметрические (∞^2) системы дифференциальных уравнений, каждое из которых имеет ∞^2 решений; полная совокупность решений должна совпадать с полным четырехпараметрическим (∞^4) семейством экстремалей. Мы доказали, что решения уравнений (33)

$$x=x(z, \alpha, \beta, A, B), \qquad y=y(z, \alpha, \beta, A, B) \tag{36}$$

образуют полное четырехпараметрическое (∞^4) семейство экстремалей; следовательно, для получения полной системы ∞^4 экстремалей из полного решения уравнения Гамильтона — Якоби $S(x, y, z, \alpha, \beta)$ необходимо только, согласно (33) и (36), продифференцировать его и привести подобные члены.

5. Канонические уравнения Гамильтона. Каждое из уравнений Эйлера (7) является дифференциальным уравнением второго порядка. Иногда оказывается удобным заменить их четырьмя дифференциальными уравнениями первого порядка. Такую замену можно провести многими способами. Наиболее симметричный способ дает так называемые канонические уравнения Гамильтона, которые выводятся следующим образом.

Соотношения (13) рассматриваются как преобразования Лежандра (см. стр. 139), заменяющие переменные u, v на U, V (сохраняя x, y, z) и функцию $F(u, v, x, y, z)$ на $W(U, V, x, y, z)$. Последнее уравнение (13) можно представить в виде

$$W = F - uU - vV, \tag{37}$$

тогда

$$dW = dF - u\,dU - v\,dV - U\,du - V\,dv.$$

Далее

$$dF = F_u\,du + F_v\,dv + F_x\,dx + F_y\,dy + F_z\,dz = U\,du + V\,dv + F_x\,dx + F_y\,dy + F_z\,dz.$$

Следовательно,

$$dW = -u\,dU - v\,dV + F_x\,dx + F_y\,dy + F_z\,dz.$$

Поскольку W нужно считать функцией U, V, x, y, z, то

$$W_U = -u, \quad W_V = -v, \quad W_x = F_x, \quad W_y = F_y, \quad W_z = F_z. \tag{38}$$

Если рассматривать теперь кривую $x = x(z)$, $y = y(z)$, которая удовлетворяет уравнениям

$$x' = u(x, y, z), \quad y' = v(x, y, z), \tag{39}$$

то эти уравнения совместно с уравнениями Эйлера (7) можно записать в виде

$$x' = -W_U, \quad y' = -W_V, \quad U' = W_x, \quad V' = W_y. \tag{40}$$

Соотношения (40) представляют собой четыре дифференциальных уравнения первого порядка относительно x, y, U, V как функций z и называются *каноническими уравнениями Гамильтона*. Их можно рассматривать как уравнения Эйлера для вариационного интеграла, выраженные через функцию $W(U, V, x, y, z)$. Если подставить (37) и (39) в (1), то вариационный интеграл запишется в виде

$$I = \int_{z_1}^{z_2} \{W(U, V, x, y, z) + x'U + y'V\}\,dz. \tag{41}$$

Если считать теперь U, V, x, y четырьмя неизвестными функциями z и написать для каждого из них уравнение Эйлера, то в результате мы получим (40).

6. Частный случай, когда независимая переменная не входит явно в подынтегральное выражение. Случай, когда F не зависит явно от z, требует отдельного рассмотрения.

В общем случае для $F(x', y', x, y, z)$ справедливо соотношение

$$\frac{dF}{dz} = F_{x'}x'' + F_{y'}y'' + F_x x' + F_y y' + F_z.$$

Предположим теперь, что $F_z = 0$, и подставим выражения для F_x и F_y из уравнений Эйлера (7). Тогда

$$\frac{dF}{dz} = F_{x'}x'' + F_{y'}y'' + x'\frac{d}{dz}F_{x'} + y'\frac{d}{dz}F_{y'} = \frac{d}{dz}(x'F_{x'} + y'F_{y'}).$$

Следовательно,

$$\frac{d}{dz}(F - x'F_{x'} - y'F_{y'}) = 0$$

т. е.

$$F - x'F_{x'} - y'F_{y'} = \text{const}. \tag{42}$$

Полученное выражение, равное значению W на экстремали, не зависит от z; следовательно, W есть постоянная интегрирования. Тот же результат можно непосредственно получить из канонических уравнений (40), так как, если F не зависит явно от z, величина W тоже не зависит от z, $W_z=0$ и

$$\frac{dW}{dz}=W_U U'+W_V V'+W_x x'+W_y y';$$

последнее же выражение обращается в нуль в соответствии с (40).

В рассматриваемом случае трехмерную вариационную задачу можно свести к двумерной. Будем считать y функцией x; тогда $y'=(dy/dx)\,x'$ и $F(x', y', x, y)=F(x', dy/dx, x, y)$. Аналогично величины $F_{x'}$, $F_{y'}$ и $F-x'F_{x'}-y'F_{y'}$ тоже можно считать функциями x', dy/dx, x, y. Решение уравнения

$$F-x'F_{x'}-y'F_{y'}=W \tag{43}$$

относительно x' имеет вид

$$x'=\Phi\left(\frac{dy}{dx}, x, y, W\right), \quad y'=\frac{dy}{dx}\Phi\left(\frac{dy}{dx}, x, y, W\right). \tag{44}$$

Рассмотрим теперь все кривые, для которых W равно некоторому заданному значению; тогда интеграл (1) можно заменить на разность двух интегралов, а именно

$$J=\int_{z_1}^{z_2}(F-W)\,dz=\int_{z_1}^{z_2}(x'F_{x'}+y'F_{y'})\,dz=\int_{x_1}^{x_2}\left(F_{x'}+\frac{dy}{dx}F_{y'}\right)dx.$$

Если ввести обозначение

$$F_{x'}+\frac{dy}{dx}F_{y'}=f \tag{45}$$

и исключить x', y' с помощью (44), то f станет функцией dy/dx, x, y, W, и мы можем написать

$$J=\int_{x_1}^{x_2} f\left(\frac{dy}{dx}, x, y, W\right)dx. \tag{46}$$

Таким образом, для каждого значения W получается вариационный интеграл с числом параметров, уменьшенным на единицу. (Такой переход соответствует в механике переходу от принципа Гамильтона к принципу Мопертюи: см. (88) ниже). Если из уравнения Эйлера, соответствующего (46), определить $y(x)$, то полное семейство экстремалей получается путем интегрирования (44).

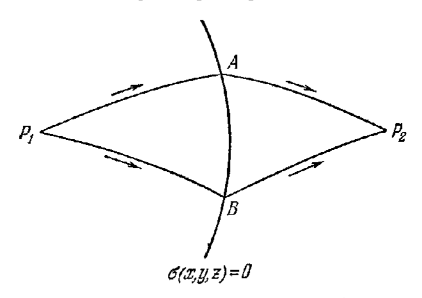


Рис. 2. Иллюстрация вариационного аналога оптического закона преломления.

7. Разрывы. Может случиться, что функция $F(u, v, x, y, z)$ не везде непрерывна. Наиболее важен случай (часто встречающийся в оптике), когда F претерпевает конечный разрыв на поверхности $\sigma(x, y, z)=0$ при всех значениях u и v.

Очевидно, что вне этой поверхности экстремали находятся как обычно из решения уравнений Эйлера (7); однако если экстремаль пересекает поверхность, ее направление резко меняется (происходит преломление). Будем различать области слева и справа от поверхности, снабжая их соответственно индексами 1 и 2 (рис. 2).

Для нахождения «закона преломления» выведем условие, при выполнении которого интеграл Гильберта S (11), взятый от точки P_1, находящейся слева от этой поверхности, до точки P_2 справа от нее, не зависит от пути интегрирования. Рассмотрим два пути P_1AP_2 и P_1BP_2, где A и B — точки на поверхности. Потребуем (обозначения очевидны), чтобы $S(P_1AP_2)=S(P_1BP_2)$ или

$$S_1(P_1A)+S_2(AP_2)=S_1(P_1B)+S_2(BP_2). \tag{47}$$

Для замкнутого контура P_1ABP_1, расположенного слева от поверхности, имеем

$$S_1(P_1A)+S_1(AB)+S_1(BP_1)=0, \tag{48}$$

а для замкнутого контура P_2BAP_2 справа от нее —

$$S_2(P_2B)+S_2(BA)+S_2(AP_2)=0. \tag{49}$$

Складывая уравнения (48) и (49) и используя (47) и соотношение $S(XY)=-S(YX)$, получим

$$S_1(AB)=S_2(AB). \tag{50}$$

Иными словами, интеграл (11), взятый по любому пути на поверхности $\sigma=0$, имеет одно и то же значение независимо от того, возьмем ли мы в качестве u, v величины u_1, v_1 слева от нашей поверхности или u_2, v_2 справа от нее. Следовательно, в обоих случаях подынтегральные выражения равны между собой, и *закон преломления* эквивалентен утверждению, что выражение

$$F+(x'-u)F_u+(y'-v)F_v \tag{51}$$

непрерывно на поверхности $\sigma=0$. Согласно (13) это условие можно записать в виде

$$(Ux'+Vy'+W)_1=(Ux'+Vy'+W)_2, \tag{52}$$

где x', y' — производные от $x(z)$, $y(z)$ для произвольной кривой на поверхности. Условие (52) соответствует также утверждению, что вектор U_2-U_1, V_2-V_1, W_2-W_1 нормален поверхности разрыва, т. е.

$$(U_2-U_1)\,dx+(V_2-V_1)\,dy+(W_2-W_1)\,dz=0. \tag{53}$$

Вывод закона отражения для экстремалей очень близок к рассмотренному выше. Для этого нужно соединить точки P_1 и P_2, расположенные по одну сторону от заданной поверхности $\sigma(x, y, z)=0$ (рис. 3) в области, где F является непрерывной функцией x, y, z, кривой P_1AP_2, которая терпит разрыв (меняет свое направление) в точке A на данной поверхности.

Рис. 3. Иллюстрация вариационного аналога оптического закона отражения.

Ясно, что отрезки P_1A и AP_2 являются экстремалями; тогда из рассуждений, сходных с теми, которые использовались при выводе закона преломления (53), следует, что условием справедливости *теоремы независимости* для падающего (индекс 1) и отраженного (индекс 2) полей является *закон отражения*

$$(U_1+U_2)\,dx+(V_1+V_2)\,dy+(W_1+W_2)\,dz=0. \tag{54}$$

Теорема независимости выполняется также для любых полей, имеющих конечное число разрывов непрерывности (как преломления, так и отражения). Как будет показано ниже, во всех таких случаях интеграл (1) имеет минимальное значение независимо от того, является ли экстремаль, соединяющая точки P_1 и P_2, непрерывной или терпит разрыв (изменяет свое направление) при условии, что функция F удовлетворяет некоторым простым условиям вдоль этой кривой.

8. Условия Вейерштрасса и Лежандра (достаточные условия экстремума). До сих пор мы не делали различия между максимумом и минимумом; рассмотренные экстремали (гладкие или с «петлями») могут даже соответствовать стационарным случаям, когда истинного экстремума вообще не существует. Выведем теперь необходимые условия истинного минимума.

Пусть $\bar{x}(z)$, $\bar{y}(z)$ — фиксированная экстремаль $\bar{C}$ в поле $u(x, y, z)$, $v(x, y, z)$, и пусть $x(z)$, $y(z)$ — любая соседняя кривая C, совпадающая с $\bar{C}$

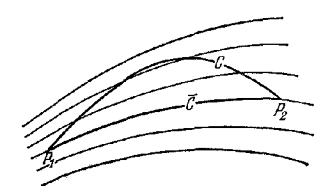


Рис. 4. К определению $\mathcal{E}$-функции Вейерштрасса.

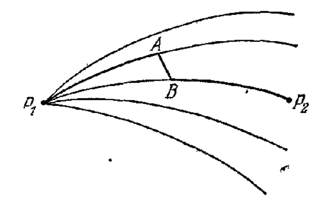


Рис 5 К выводу условия сильного минимума Вейерштрасса

в концевых точках P_1 и P_2 и тоже целиком лежащая в поле u, v (рис. 4). Экстремум является истинным минимумом, если

$$\int_C F(x', y', x, y, z)\,dz - \int_{\bar{C}} F(\bar{x}', \bar{y}', \bar{x}, \bar{y}, z)\,dz > 0. \tag{55}$$

Согласно пп. 2 и 3 второй интеграл в последнем соотношении можно заменить на интеграл вдоль кривой C, т. е. на

$$\int_C \{F(u, v, x, y, z) + (x' - u) F_u + (y' - v) F_v\}\,dz;$$

этот интеграл не зависит от пути и сводится к $\int_{\bar{C}} F\,dz$, если путь интегрирования совпадает с $\bar{C}$. Следовательно, (55) принимает вид

$$\int_C F\,dz - \int_{\bar{C}} F\,dz = \int_C \mathcal{E}(x', y', u, v, x, y, z)\,dz > 0, \tag{56}$$

где

$$\mathcal{E}(x', y', u, v, x, y, z) = \\ = F(x', y', x, y, z) - F(u, v, x, y, z) - (x' - u) F_u - (y' - v) F_v, \tag{57}$$

а аргументы функций F_u, F_v совпадают с аргументами функции $F(u, v, x, y, z)$. Функция, определенная (57), называется *$\mathcal{E}$-функцией* (*или функцией избытка*) *Вейерштрасса*; аргументы x, y, z, x', y' характеризуют точку на кривой C и направление последней, тогда как u, v — направление экстремали поля, проходящей через точку x, y, z.

Как мы видим, $\mathcal{E}$ обращается в нуль на любом отрезке кривой C, совпадающем с экстремалью поля. Выберем теперь в качестве поля двухпараметрическое (∞^2) семейство всех экстремалей, проходящих через точку P_1. Затем проведем специальную кривую C так, чтобы она совпадала с экстремалью поля между точками P_1 и A, являлась прямой между точкой A и точкой B, лежащей на заданной экстремали, и совпадала с заданной экстремалью между B и P_2 (рис. 5). Тогда $\mathcal{E}$ на отрезках P_1A и BP_2 обращается в нуль и (56) принимает вид

$$\int_A^B \mathcal{E}\,dz > 0,$$

Устремляя A к B, получим, что это неравенство выполняется только в том случае, если

$$\mathcal{E}(x', y', \bar{x}', \bar{y}', \bar{x}, \bar{y}, \bar{z}) > 0, \qquad (58)$$

где $\bar{x}$, $\bar{y}$, $\bar{z}$ — координаты произвольной точки (B) на данной экстремали $\bar{C}$, а x', y' относятся к совершенно произвольному направлению AB. Формула (58) служит *условием сильного минимума Вейерштрасса*; разумеется, оно является необходимым условием. Однако если предположить, что функция F непрерывна по всем своим пяти переменным (следовательно, функция $\mathcal{E}$ непрерывна по своим семи переменным), то при условии, что (58) удовлетворяется для всех точек заданной экстремали и при произвольном направлении x', y', неравенство (56) должно выполняться для любой соседней кривой C, имеющей произвольное направление и лежащей в некоторой окрестности кривой $\bar{C}$. Следовательно, условие (58) является также достаточным условием сильного минимума. Это, конечно, относительный минимум, так как может найтись несколько экстремалей, для которых интеграл (1) минимален по сравнению с интегралом по всем соседним кривым; рассмотренным способом нельзя установить, какой из этих минимумов абсолютный.

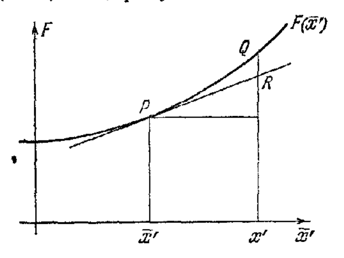


Рис. 6. Геометрическое истолкование условия сильного минимума Вейерштрасса.

Неравенство (58) допускает простую геометрическую интерпретацию. Для фиксированной точки пространства $\bar{x}$, $\bar{y}$, $\bar{z}$ величина F зависит только от $\bar{x}'$, $\bar{y}'$; эту функцию $F(\bar{x}', \bar{y}')$ можно представить поверхностью в трехмерном пространстве $\bar{x}'$, $\bar{y}'$, F. (На рис. 6 изображен двумерный разрез $\bar{x}'F$). Тогда

$$\mathcal{E}(x', y', \bar{x}', \bar{y}') = F(x', y') - [F(\bar{x}', \bar{y}') + (x' - \bar{x}')F_{\bar{x}'} + (y' - \bar{y}')F_{\bar{y}'}] \qquad (59)$$

есть, очевидно, расстояние QR вдоль ординаты в пространстве x', y' между точкой Q на поверхности $F = F(\bar{x}', \bar{y}')$ и точкой R (см. рис. 6), в которой плоскость, касающаяся в точке $P(\bar{x}', \bar{y}')$ поверхности F, пересекает эту ординату. Следовательно, $\mathcal{E}(x', y', \bar{x}', \bar{y}') > 0$, когда поверхность F находится над этой касательной плоскостью. Если это справедливо для всех x', y', то имеется *сильный минимум*.

Если же условие (58) выполняется лишь в небольшой области $\xi = x' - \bar{x}'$, $\eta = y' - \bar{y}'$, то имеется слабый минимум; в этом случае величину $\mathcal{E}$ можно разложить по степеням ξ, η, и тогда (снова опуская аргументы x, y, z) получим

$$\mathcal{E}(x', y', \bar{x}', \bar{y}') =$$
$$= F(x', y') - F(\bar{x}', \bar{y}') - \xi F_{\bar{x}'} - \eta F_{\bar{y}'} + \frac{1}{2}[F_{\bar{x}'\bar{x}'}\xi^2 + 2F_{\bar{x}'\bar{y}'}\xi\eta + F_{\bar{y}'\bar{y}'}\eta^2] + \ldots$$

При малых значениях ξ и η квадратичные члены играют решающую роль, и для существования минимума они должны быть положительными. Таким образом, *условие Лежандра* (необходимое и достаточное) *для слабого минимума* *) имеет вид

$$F_{\bar{x}'\bar{x}'} > 0, \qquad F_{\bar{x}'\bar{x}'}F_{\bar{y}'\bar{y}'} - F^2_{\bar{x}'\bar{y}'} > 0. \qquad (60)$$

*) Это условие легко обобщить на случай большего числа переменных (скажем, n). Для существования минимума квадратичная форма с n переменными должна быть положительно определенной; отсюда следует, что соответствующий определитель и все его главные миноры должны быть положительными.

9. *Минимум вариационного интеграла, когда один конец кривой связан с поверхностью.* С помощью $\mathscr{E}$-функции легко определить минимум вариационного интеграла (1) относительно всех кривых с одной общей концевой точкой P_1, а другими, связанными с заданной поверхностью $\sigma(x, y, z)=0$.

Кривая, очевидно, должна совпадать с одной из экстремалей двухпараметрического ∞^2 семейства, проходящей через точку P_1, и нужно лишь узнать, «с какой из них». Среди всех ∞^2 экстремалей имеется только одна *), которая ортогональна поверхности $\sigma=0$, и легко показать, что именно она служит решением данной задачи. Для этого обозначим через P_2 точку пересечения такой экстремали с поверхностью $\sigma=0$ и окружим ее полем всех экстремалей, перпендикулярных к поверхности. Пусть P_1Q — любая экстремаль, проходящая через P_1, а Q — точка ее пересечения с поверхностью (рис. 7). Тогда интеграл Гильберта $S(P_2, Q)$ обращается в нуль. Следовательно, интеграл S вдоль пути P_1QP_2 равен вариационному интегралу $I(P_1, Q)$. Разность $I(P_1, Q) - I(P_1, P_2)$ можно выразить точно таким же образом, как и выше, через $\mathscr{E}$-функцию; тогда, если условие (58) выполняется, она положительна для всех Q, отличных от P_1.

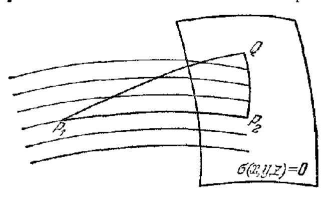


Рис. 7. К определению минимума вариационного интеграла для всех кривых, один конец которых фиксирован, а другой связан с поверхностью.

10. Критерий Якоби для минимума. Если экстремаль можно целиком поместить в поле и условие Лежандра выполняется для всех ее точек между P_1 и P_2, то интеграл I, определенный (1), имеет (слабый) минимум. Остается найти критерий существования такого поля.

Пусть двухпараметрическое (∞^2) семейство экстремалей, проходящих через P_1, описывается уравнениями

$$x=x(z, \alpha, \beta), \qquad y=y(z, \alpha, \beta), \tag{61}$$

и пусть заданная экстремаль C характеризуется $\alpha=0$, $\beta=0$, т. е.

$$x=x(z, 0, 0), \qquad y=y(z, 0, 0). \tag{62}$$

Кривые (61) образуют поле, если существует одна кривая, проходящая через заданную точку $P(x, y)$ на произвольно близком расстоянии от C, т. е. если уравнения (61) имеют единственное решение для α, β как функций x, y; это условие имеет вид

$$\Delta=\begin{vmatrix} x_\alpha & x_\beta \\ y_\alpha & y_\beta \end{vmatrix} \neq 0. \tag{63}$$

Полученное соотношение является *критерием Якоби существования минимума.*

Определитель Δ есть функция z вдоль заданной экстремали (62). Первая точка $\bar{P}$, в которой Δ обращается в нуль, называется точкой, *сопряженной точке* P_1; для любого интервала P_1P_2, где P_2 лежит между P_1 и $\bar{P}$, существует истинный минимум.

В точке $\bar{P}$ заданную экстремаль пересекает соседняя (бесконечно близкая к заданной); эта точка лежит на огибающей семейства (61). Следовательно, *поле ограничено огибающей семейства экстремалей* (61). В оптике такие огибающие являются каустическими поверхностями.

11. Пример I. Оптика. Проиллюстрируем теперь общую теорию несколькими примерами. В первом примере рассматриваются наикратчайшая линия в обычной геометрии и кратчайший оптический путь в геометрической оптике.

*) Предполагается, что P_1 находится достаточно близко к поверхности, т. е. случай нескольких таких экстремалей исключен.

Геометрия Евклида основана на теореме Пифагора, согласно которой элемент длины ds связан со своими проекциями dx, dy, dz на оси прямоугольной системы координат соотношением

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{64}$$

Тогда кратчайшие линии между двумя точками определяются из минимума вариационного интеграла

$$s = \int_{P_1}^{P_2} ds = \int_{z_1}^{z_2} \sqrt{x'^2 + y'^2 + 1}\, dz. \tag{65}$$

Геометрическую оптику можно построить с помощью обобщения этого интеграла, а именно с помощью принципа наикратчайшего оптического пути Ферма (см. п. 3.3.2), т. е.

$$\int_{P_1}^{P_2} n\, ds = \int_{z_1}^{z_2} n(x, y, z) \sqrt{x'^2 + y'^2 + 1}\, dz, \tag{66}$$

где $n(x, y, z)$ — показатель преломления. Мы будем рассматривать только второй случай, так как соотношение (65) является частным случаем (66) при $n = 1$.

Имеем

$$F(x', y', x, y, z) = n(x, y, z) \sqrt{x'^2 + y'^2 + 1}. \tag{67}$$

Поскольку $ds = \sqrt{x'^2 + y'^2 + 1}\, dz$, имеем

$$\left.\begin{aligned} U &= F_{x'} = \frac{nx'}{\sqrt{x'^2 + y'^2 + 1}} = n\frac{dx}{ds} = ns_x,\\ V &= F_{y'} = \frac{ny'}{\sqrt{x'^2 + y'^2 + 1}} = n\frac{dy}{ds} = ns_y,\\ W &= F - F_{x'}x' - F_{y'}y' = \frac{n}{\sqrt{x'^2 + y'^2 + 1}} = n\frac{dz}{ds} = ns_z, \end{aligned}\right\} \tag{68}$$

где s_x, s_y, s_z — компоненты единичного вектора $\mathbf{s}$, касательного к кривой $x = x(z)$, $y = y(z)$. Закон преломления на поверхности разрыва $n(x, y, z)$ можно, согласно (53), записать в виде

$$(n_2\mathbf{s}_2 - n_1\mathbf{s}_1)\cdot d\mathbf{l} = 0, \tag{69а}$$

где $d\mathbf{l}$ (dx, dy, dz) — произвольный элемент длины на поверхности. Из последнего соотношения следует, что векторы $\mathbf{s}_1$, $\mathbf{s}_2$ и нормаль к поверхности разрыва компланарны и что углы θ_1 и θ_2, образованные $\mathbf{s}_1$ и $\mathbf{s}_2$ с нормалью к поверхности соответственно, связаны соотношением

$$n_2 \sin\theta_2 = n_1 \sin\theta_1, \tag{69б}$$

что согласуется с законом преломления (3.2.19).

Уравнения Эйлера (7), соответствующие (66), имеют вид

$$\frac{\partial n}{\partial x}\sqrt{x'^2 + y'^2 + 1} - \frac{d}{dz}\frac{nx'}{\sqrt{x'^2 + y'^2 + 1}} = 0,$$

$$\frac{\partial n}{\partial y}\sqrt{x'^2 + y'^2 + 1} - \frac{d}{dz}\frac{ny'}{\sqrt{x'^2 + y'^2 + 1}} = 0,$$

или

$$\frac{d}{ds}\left(n\frac{dx}{ds}\right) = \frac{\partial n}{\partial x}, \qquad \frac{d}{ds}\left(n\frac{dy}{ds}\right) = \frac{\partial n}{\partial y}. \tag{70а}$$

Третье уравнение Эйлера

$$\frac{d}{ds}\left(n\frac{dz}{ds}\right) = \frac{\partial n}{\partial z} \tag{70б}$$

является тождеством, так как оно следует из соотношения

$$\left(\frac{dx}{ds}\right)^2+\left(\frac{dy}{ds}\right)^2+\left(\frac{dz}{ds}\right)^2=1. \qquad (71)$$

Для получения (70б) можно поступить следующим образом: сначала продифференцируем (71) по s. Затем умножим (71) на dn/ds, а продифференцированное уравнение — на n, после чего сложим полученные выражения. И, наконец, используем *) (70а). Три скалярных дифференциальных уравнения (70) согласуются с векторным уравнением (3.2.2) для световых лучей.

Поскольку U, V и W характеризуют теперь компоненты лучевого вектора (см. п. 4.1.4), из (12) следует, что функция S, соответствующая полю геометрической оптики, является *точечной характеристикой Гамильтона* (см. (4.1.1)). Более того, с помощью процедуры, использовавшейся при выводе (21), можно показать, что уравнение Гамильтона — Якоби для данной вариационной задачи совпадает с *уравнением эйконала*.

Для изучения условия Лежандра (60), необходимо найти производные $F_{x'x'}$ и т. д.

В данном случае имеем

$$F_{x'x'}=n\frac{1+y'^2}{(1+x'^2+y'^2)^{3/2}},$$

$$F_{y'y'}=n\frac{1+x'^2}{(1+x'^2+y'^2)^{3/2}},$$

$$F_{x'y'}=-n\frac{x'y'}{(1+x'^2+y'^2)^{3/2}};$$

следовательно,

$$F_{x'x'}F_{y'y'}-F_{x'y'}^2=\frac{n^2}{(1+x'^2+y'^2)^2}>0. \qquad (72)$$

Таким образом, каждой экстремали соответствует слабый минимум (см. стр. 674), если выполняется условие Якоби (63). Однако, поскольку выпуклость функции F для данных значений x, y, z, т. е. для данного n, направлена вниз при всех x', y', то из геометрического истолкования условия Вейерштрасса следует, что минимум является сильным.

Остается рассмотреть критерий Якоби. При $n=\text{const}$, т. е. в обычной евклидовой геометрии, экстремали, очевидно, имеют вид прямых. Так как пучок прямых, проходящих через точку P_1, не может иметь огибающую, то каждой прямой линии соответствует сильный минимум расстояния между любыми двумя точками на этой прямой. В геометрической же оптике, когда n является в общем случае функцией x, y, z (непрерывной или разрывной), пучок лучей из точки P_1 может иметь огибающую (каустическую поверхность). Поэтому при исследовании характера экстремума необходимо в каждом случае отдельно рассматривать такие поверхности.

12. Пример II. Механика материальных точек. В качестве второго примера рассмотрим *механику системы материальных точек*. Независимой переменной здесь выступает время t, а в качестве неизвестных функций — лагранжевы координаты q_α ($\alpha=1, 2, \ldots, n$) и их производные, т. е. скорости, $u_\alpha=\dot{q}_\alpha$.

Вариационный характер задачи определяется *принципом Гамильтона*

$$\int_{t_1}^{t_2} L(u_1, u_2, \ldots, q_1, q_2, \ldots, t)\,dt=\text{экстремум}, \qquad (73)$$

*) Уравнение (70б) можно получить более симметричным способом, однако для этого придется использовать лишнее уравнение Эйлера. Координаты x, y, z считаются функциями параметра λ; тогда получаются три уравнения Эйлера, связанные между собой тождеством, а параметр λ идентифицируется в конце концов с переменной s.

где L — *лагранжиан*. В обычной нерелятивистской механике $L=T-\Phi$, где T — кинетическая энергия, т. е. квадратичная форма по u_α, и Φ — потенциальная энергия; однако условие (73) справедливо и в более общем случае, когда действует магнитная сила и учитывается релятивистское изменение массы.

Функции $F(u, v, x, y, z)$ соответствует теперь $L(u, q, t)$, и следовательно, (см. (13)) величинам U, V — импульсы

$$p_\alpha = \frac{\partial L}{\partial u_\alpha}, \tag{74}$$

а величине W соответствует $-H$, где H — *гамильтониан* вида

$$H = \sum_\alpha u_\alpha \frac{\partial L}{\partial u_\alpha} - L = \sum_\alpha u_\alpha p_\alpha - L. \tag{75}$$

Если $L=T-\Phi$, то последнее соотношение принимает вид

$$H = \sum_\alpha u_\alpha \frac{\partial L}{\partial u_\alpha} - T + \Phi.$$

Из теоремы Эйлера для однородных функций *) следует, что

$$2T = \sum u_\alpha \frac{\partial T}{\partial u_\alpha} \left(= \sum u_\alpha \frac{\partial L}{\partial u_\alpha}\right),$$

и H сводится к выражению для полной энергии, т. е.

$$H = T + \Phi. \tag{76}$$

Уравнения Эйлера (7) дают уравнения движения Лагранжа

$$\frac{d}{dt}\left(\frac{\partial L}{\partial u_\alpha}\right) - \frac{\partial L}{\partial q_\alpha} = 0, \tag{77}$$

и канонические уравнения (40) принимают вид

$$\frac{dq_\alpha}{dt} = \frac{\partial H}{\partial p_\alpha}, \qquad \frac{dp_\alpha}{dt} = -\frac{\partial H}{\partial q_\alpha}, \tag{78}$$

где H считается функцией p_α и q_α. Если H не зависит от времени t, то формула (42) выражает закон сохранения энергии

$$H = \sum_\alpha u_\alpha p_\alpha - L = \text{const} = E. \tag{79}$$

В этом случае дифференциальное уравнение Гамильтона — Якоби (21) записывается следующим образом:

$$\frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial q_\alpha}, q_\alpha\right) = 0. \tag{80}$$

Интегрируя его, получим

$$S = -Et + S_1(q_\alpha), \tag{81}$$

где S_1 удовлетворяет уравнению

$$H\left(\frac{\partial S_1}{\partial q_\alpha}, q_\alpha\right) = E. \tag{82}$$

Из решения уравнения Гамильтона — Якоби получаем в согласии с (17) выражение для импульсов в виде

$$p_\alpha = \frac{\partial S}{\partial q_\alpha} = \frac{\partial S_1}{\partial q_\alpha}. \tag{83}$$

Отсюда следует, что линейный интеграл

$$\int_{P_1}^{P_2} \sum_\alpha p_\alpha dq_\alpha \tag{84}$$

---

*) См., например, [1].

не зависит от пути, соединяющего P_1 и P_2, и поэтому он равен нулю, если путь интегрирования — замкнутый в односвязной области, т. е.

$$\oint p_\alpha dq_\alpha = 0. \tag{85}$$

Если функции p_α многозначны, то интеграл по замкнутому контуру может не обратиться в нуль, а равняться числу, кратному какому-нибудь постоянному периоду. Найденное соотношение является обобщением оптического инварианта Лагранжа (см. п. 3.3.1) и совпадает с одним из *инвариантов Пуанкаре*. Его можно также представить в виде

$$\frac{\partial p_\alpha}{\partial q_\beta} - \frac{\partial p_\beta}{\partial q_\alpha} = 0. \tag{86}$$

Если H не зависит от времени, то переменную t можно исключить из выражения для принципа минимума с помощью формул (45) и (46). В результате получим

$$J = \int_{t_1}^{t_2} (L+E)\,dt = \int_{t_1}^{t_2} \sum_\alpha \dot{q}_\alpha p_\alpha\,dt = \int_{P_1}^{P_2} \sum_\alpha p_\alpha\,dq_\alpha = \text{экстремум}. \tag{87}$$

Это есть *принцип наименьшего действия Мопертюи*, обобщенный для произвольного лагранжиана L; его нужно понимать следующим образом: уравнение (79) позволяет исключить производные по времени $u_\alpha = \dot{q}_\alpha$, выражая их через производные, скажем dq_α/dq_1 ($\alpha = 2, \ldots, n$), если считать q_1 независимой переменной. Уравнение (87) выражает чисто геометрический принцип, который описывает траектории, но не движения. Последние можно найти из соотношения (78).

Если $L = T - \Phi$, то $E = T + \Phi$, и мы получим первоначальную формулу Мопертюи

$$J = 2\int_{t_1}^{t_2} T\,dt, \tag{88}$$

которую следует трактовать таким же образом. Рассмотрение вопросов электронной оптики (см. п. 2 приложения 2) служит примером подобного сведения задачи о движении к задаче о траекториях посредством перехода от принципа Гамильтона к принципу Мопертюи.

Условие Лежандра можно рассматривать только в случае заданного L. Если $L = T - \Phi$ и T — квадратичная форма по u_α, то это условие, очевидно, эквивалентно утверждению, что T — положительно определенная величина. Тогда условие Вейерштрасса тоже выполняется, и имеется сильный минимум, если удовлетворяется условие Якоби. Последнее позволяет исследовать динамические фокусы и каустики, что не представляет большого практического значения.

Условие Лежандра для релятивистского электрона с лагранжианом

$$L = -mc^2\sqrt{1-\left(\frac{v}{c}\right)^2} + e\left(\frac{\mathbf{v}}{c}\cdot\mathbf{A} - \varphi\right) \tag{89}$$

(см. приложение 2) дает форму

$$\frac{m}{\left[1-\left(\frac{v}{c}\right)^2\right]^{3/2}}\left\{\rho^2\left(1-\frac{v^2}{c^2}\right)+\left(\frac{\boldsymbol{\rho}\cdot\mathbf{v}}{c}\right)^2\right\}, \tag{90}$$

квадратичную по компонентам вектора $\boldsymbol{\rho}(\xi, \eta, \zeta)$, и поэтому оно всегда удовлетворяется.

ПРИЛОЖЕНИЕ 2

ОБЫЧНАЯ ОПТИКА, ЭЛЕКТРОННАЯ ОПТИКА И ВОЛНОВАЯ МЕХАНИКА

В 1831 г. Вильям Гамильтон обнаружил аналогию между траекториями материальных частиц в потенциальных полях и траекториями световых лучей в средах с непрерывно изменяющимся показателем преломления. Благодаря своему математическому изяществу «аналогия Гамильтона» излагалась в течение почти ста лет в учебниках по динамике, но практически ее никто не применял до 1925 г., когда Буш впервые объяснил фокусирующее действие электрических и магнитных полей на электронные пучки, пользуясь оптической терминологией. Примерно в то же время Шредингер воспользовался аналогией Гамильтона для получения своего уравнения, позволившего перейти от геометрической оптики к волновой оптике частиц; при этом он использовал понятие длины волны частиц, впервые предложенное в 1923 г. де Бройлем.

Практическая электронная оптика начала бурно развиваться с 1928 г. В это время аналогия Гамильтона была уже широко известна и ее использование позволило изобрести целый ряд электроннооптических приборов (таких, как электронный микроскоп), являющихся аналогами оптических. Хотя математическая аналогия носит общий характер, оптическая и электронная техники различаются между собой. Такие электроннооптические приборы, как электроннолучевые трубки и системы с искривленной оптической осью, не имеют аналогов в обычной оптике. Мы будем рассматривать только такие проблемы электронной оптики, оптические аналоги которых подробно обсуждались в предыдущих главах книги, и поэтому получающиеся результаты после небольшой модификации почти *полностью* переходят в уже известные. Отметим, что это относится, в частности, к наиболее трудной для понимания главе электронной оптики, а именно к волновой теории аберраций линз.

1. Аналогия Гамильтона в элементарной форме. Мы сначала покажем, что задачу определения траектории заряженной частицы можно свести к оптической задаче путем введения подходящего показателя преломления, изменяющегося от точки к точке.

Рассмотрим частицу с зарядом e и массой m и для простоты будем считать ее электроном, движущимся в постоянном электростатическом поле с потенциалом $\varphi(x, y, z)$. Из уравнения движения Ньютона имеем

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} = -e \operatorname{grad} \varphi, \tag{1}$$

где $\mathbf{p}$ — вектор импульса. Это уравнение справедливо для любых скоростей $\mathbf{v}$, если ньютоновское определение импульса $\mathbf{p} = m\mathbf{v}$ заменить на эйнштейновское

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-\beta^2}}, \qquad \beta = \frac{v}{c}, \tag{2}$$

где c — скорость света в вакууме.

Уравнение движения (1) удобно разбить на два; первое является уравнением траектории, а второе характеризует «временно́е расписание», согласно которому электрон движется вдоль траектории. Для этого напишем $\mathbf{v} = v\mathbf{s}$, $\mathbf{p} = p\mathbf{s}$, где $\mathbf{s}$ — единичный вектор в направлении движения. Тогда

$$\frac{d\mathbf{p}}{dt} = \frac{dp}{dt}\mathbf{s} + p\frac{d\mathbf{s}}{dt} = \frac{dp}{dt}\mathbf{s} + p\frac{d\mathbf{s}}{ds}\frac{ds}{dt} = \frac{dp}{dt}\mathbf{s} + pv\frac{d\mathbf{s}}{ds}.$$

Из дифференциальной геометрии известно, что вектор $d\mathbf{s}/ds$ направлен вдоль главной нормали $\boldsymbol{\nu}$, а его абсолютная величина равна кривизне $1/\rho$ траектории. Следовательно,

$$\frac{d\mathbf{p}}{dt} = \frac{dp}{dt}\mathbf{s} + \frac{pv}{\rho}\boldsymbol{\nu},$$

Отсюда и из уравнения (1) следует, что мгновенный центр кривизны лежит в плоскости, проходящей через касательную s к траектории и вектор электрического поля $\mathbf{E} = -\operatorname{grad}\varphi$. Разлагая $\operatorname{grad}\varphi$ по двум направлениям, имеем

$$\frac{dp}{dt}\mathbf{s} + \frac{pv}{\rho}\mathbf{v} = -e\left[(\mathbf{s}\cdot\operatorname{grad}\varphi)\,\mathbf{s} + (\mathbf{v}\cdot\operatorname{grad}\varphi)\,\mathbf{v}\right]. \tag{3}$$

Приравнивая первые члены в левой и правой частях (3), получим скалярное уравнение, которое можно назвать «временны́м расписанием», поскольку оно определяет положение электрона на траектории в зависимости от времени. После умножения на $v = ds/dt$ его можно проинтегрировать, что дает

$$\frac{mc^2}{\sqrt{1-\beta^2}} = -e\varphi + \text{const}. \tag{4}$$

Это есть эйнштейновский интеграл энергии. Для медленно движущихся частиц ($\beta \ll 1$) он переходит в ньютоновский интеграл $\frac{1}{2}mv^2 + e\varphi = \text{const}$.

Ограничимся теперь для удобства рассмотрением электронов, для которых постоянные интегрирования одинаковы, т. е. обладающих одинаковой полной энергией. Этот случай соответствует электронам, вылетающим из определенной потенциальной поверхности φ_0 с нулевой скоростью. Во многих практических задачах такая поверхность совпадает с поверхностью катода. Полагая

$$V = \varphi - \varphi_0,$$

т. е. считая поверхность φ_0 началом отсчета потенциала V, интеграл энергии можно представить в виде

$$mc^2\left[\frac{1}{\sqrt{1-\beta^2}} - 1\right] = -eV. \tag{4a}$$

Комбинируя (2) и (4a), можно написать одно двойное уравнение

$$\left(1 - \frac{eV}{mc^2}\right)^2 = 1 + \left(\frac{p}{mc}\right)^2 = \frac{1}{1-\beta^2}. \tag{5}$$

Тогда абсолютная величина импульса p этих частиц, выраженная через координаты x, y, z, примет вид

$$mc^2\left[\sqrt{1 + \left(\frac{p}{mc}\right)^2} - 1\right] = -eV(x, y, z). \tag{6}$$

Рассмотрим теперь вторую часть (3), т. е. компоненту, перпендикулярную к направлению движения,

$$\frac{pv}{\rho} = -e\left[\mathbf{v}\cdot\operatorname{grad} V\right]. \tag{7}$$

Выражая v и V с помощью (2) и (6) через p, получим простой закон

$$\frac{1}{\rho} = \frac{\mathbf{v}\cdot\operatorname{grad} p}{p} = \mathbf{v}\cdot\operatorname{grad}(\ln p). \tag{8}$$

Уравнение (8) *идентично уравнению* (3.2.14) *для кривизны лучей в среде с показателем преломления* n, если считать, что последний пропорционален p; таким образом, мы получаем формальную аналогию между траекториями лучей и электронов.

Необходимо подчеркнуть, что абсолютная величина импульса p зависит только от координат лишь для электронов с фиксированной полной энергией; для электронов с различными энергиями значение p будет другой функцией, определяемой (6). Таким образом, показатель преломления зависит от энергии электронов. Это также имеет аналогию в обычной оптике, где показатель преломления среды зависит от длины волны света. Позднее будет показано существование довольно глубокой аналогии, поскольку в обоих случаях показатель преломления оказывается функцией длины волны.

Для медленных электронов значение p пропорционально скорости, а последняя в свою очередь пропорциональна $\sqrt{V}$. Использовавшиеся нами реля-

тивистские уравнения имеют следующее преимущество: они ясно показывают, что характеристической величиной служит не скорость, а импульс. Более того, найденные результаты допускают немедленное обобщение на случай общего статического поля, электрического и магнитного. Из теории относительности хорошо известно, что при наличии магнитного поля механический импульс $\mathbf{p}$, который будет теперь обозначаться через $\mathbf{p}_m$, заменяется на «полный» импульс

$$\mathbf{p}_{\text{полн}} = \mathbf{p}_m + e\mathbf{A}, \tag{9}$$

где $\mathbf{A}$ — векторный потенциал *). Отсюда можно заключить, что в общем случае показатель преломления, который в электростатике пропорционален компоненте $\mathbf{p}_m$ в направлении движения, должен заменяться на компоненту $\mathbf{p}_{\text{полн}}$ в том же направлении. Такое предположение оказывается справедливым, однако при рассмотрении законов электронной оптики при наличии электромагнитных полей мы все же прибегнем к более строгому и общему обоснованию.

2. **Аналогия Гамильтона в вариационной форме.** Законы геометрической оптики можно вывести из принципа Ферма (см. п. 3.3.2), согласно которому путь света между двумя какими-нибудь точками P_1 и P_2 обеспечивает минимум оптической длины, т. е.

$$\int_{P_1}^{P_2} n\,ds = \text{минимум}. \tag{10}$$

Необходимо напомнить, что такая *строгая* формулировка принципа Ферма справедлива лишь в том случае, когда две концевые точки расположены достаточно близко друг от друга или, что то же самое, когда на отрезке луча, соединяющего эти точки, нет изображения ни точки P_1, ни точки P_2. Если P_2 служит изображением P_1, то уравнение (10) определяет не луч, а бесконечно тонкий пучок лучей, соединяющих эти две точки и проходящих равные оптические пути. Если расстояние между P_1 и P_2 настолько велико, что между ними имеется изображение, то луч определяется из *слабой* формулировки принципа Ферма

$$\delta \int_{P_1}^{P_2} n\,ds = 0; \tag{10a}$$

такой формулировке соответствует не максимальное, как иногда ошибочно утверждают, а стационарное значение интеграла, которое не является ни максимальным, ни минимальным.

Некоторые следствия из принципа Ферма обсуждались в п. 11 приложения 1 при вычислении вариаций. В п. 12 того же приложения было показано, что движение системы материальных точек также можно записать в вариационной форме, используя принцип Гамильтона, который в частном случае одной материальной точки имеет вид

$$\int_{P_1,\,t_1}^{P_2,\,t_2} L(\dot{x}, \dot{y}, \dot{z}, x, y, z)\,dt = \text{минимум}. \tag{11}$$

Такая формулировка справедлива в любой системе координат, но мы ограничимся для простоты декартовыми координатами x, y, z. Здесь $\dot{x}$, $\dot{y}$, $\dot{z}$ — компоненты скорости, L — лагранжиан. Положение начальной и конечной точек рассматривается в четырехмерном пространстве — времени и считается неизменным в вариационном процессе. В приложении 1 (см. (77) и (78)) было показано, как из (11) получить уравнения движения в форме Лагранжа и в форме Гамильтона.

*) Необходимо отметить, что этот фундаментальный результат был получен Шварцшильдом [2] за три года до появления теории относительности.

Для релятивистского электрона с зарядом e и массой покоя m лагранжиан имеет вид

$$L = -mc^2\sqrt{1-\beta^2} - e\left(\varphi - \frac{1}{c}\dot{\mathbf{r}}\cdot\mathbf{A}\right). \quad (12)$$

Здесь $\dot{\mathbf{r}} = \mathbf{v}$ — вектор скорости, φ — электростатический потенциал, $\mathbf{A}$ — магнитный или векторный потенциал *). Лагранжиан (12) можно проверить, если подставить его в уравнения движения Лагранжа (см. (77) приложения 1) и использовать электромагнитные соотношения (2.1.5) и (2.1.7)

$$\mathbf{E} = -\frac{1}{c}\dot{\mathbf{A}} - \operatorname{grad}\varphi, \quad \mathbf{B} = \operatorname{rot}\mathbf{A}.$$

Уравнения Лагранжа принимают тогда вид

$$\frac{d\mathbf{p}}{dt} = e\left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right), \quad (13)$$

т. е. совпадают с уравнениями движения в форме Ньютона — Лоренца.

Принцип Гамильтона (11) является слишком общим для электронной оптики. Во-первых, в соответствующую вариационную задачу входит время, которое не представляет никакого интереса, если поля стационарны. Во-вторых, решение этой задачи содержит пятипараметрическое (∞^5) семейство экстремалей; четыре параметра обусловлены тем, что при данном значении полной энергии точки P_1, P_2 можно свободно выбирать на любых двух заданных поверхностях, а пятый тем, что значение полной энергии считается произвольным. В электронной оптике, как и в обычной оптике, стремятся по возможности уменьшить число измерений.

Рассмотрим сначала фиксированную энергию, другими словами, ограничимся обсуждением монохроматического света или моноэнергетических электронов. Как следует из (87) приложения 1, число измерений задачи уменьшается при этом на единицу. Принцип Гамильтона (11) заменяется принципом наименьшего действия

$$J = \int_{P_1, t_1}^{P_2, t_2} (L + E)\,dt = \text{минимум}, \quad (14)$$

где E — полная энергия, т. е.

$$E = \sum \dot{x} p_x - L.$$

Время t входит в соотношение (14) лишь формально, так как после подстановки в него выражения для E получим

$$J = \int_{P_1, t_1}^{P_2, t_2} \sum p_x \dot{x}\,dt = \int_{P_1}^{P_2} \sum p_x\,dx = \int_{P_1}^{P_2} \mathbf{p}\cdot d\mathbf{r} = \text{минимум}. \quad (15)$$

Во втором и третьем выражениях опущены t_1 и t_2, так как мы рассматриваем только постоянные во времени поля (в противном случае полная энергия не оставалась бы постоянной), в которых начальный момент t_1 не играет роли, а время $t_2 - t_1$ полностью определяется траекторией и величиной энергии.

Таким образом, принцип наименьшего действия (14) полностью аналогичен принципу Ферма (10). Изучение движения электрона сводится к оптической задаче, если определить *электронно-оптический показатель преломления как компоненту импульса в направлении движения*. Для чисто электростатического поля этот результат эквивалентен результату, полученному выше из более простых соображений. В этом случае импульс оказывается чисто механическим и параллельным траектории, а его абсолютная величина дается

*) Нерелятивистская функция Лагранжа (12) была получена Шварцшильдом [2] в 1903 г.

выражением (2), т. е.

$$p_m = \frac{mv}{\sqrt{1-\beta^2}}.$$

При наличии магнитного поля необходимо использовать общее определение компонент импульса (см. соотношение (74) приложения 1) как производных от лагранжиана по компонентам скорости. Для одной частицы с лагранжианом (12) получим

$$p_x = \frac{\partial L}{\partial \dot{x}} = \frac{m\dot{x}}{\sqrt{1-\beta^2}} + \frac{e}{c} A_x$$

и т. д. В векторном виде имеем

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1-\beta^2}} + \frac{e}{c}\mathbf{A}. \qquad (16)$$

Таким образом, общее выражение для электроннооптического показателя преломления имеет, с точностью до произвольного постоянного множителя, следующий вид:

$$n = \frac{mv}{\sqrt{1-\beta^2}} + \frac{e}{c}\mathbf{A}\cdot\mathbf{s}, \qquad (17)$$

где $\mathbf{A}\cdot\mathbf{s}$ — компонента векторного потенциала в направлении движения. Необходимо еще раз подчеркнуть, что эту величину следует рассматривать как функцию положения электронов с заданной полной энергией.

По-видимому, имеется существенное различие между общим случаем и частным случаем чисто электрических полей. В электрическом поле показатель преломления пропорционален механическому импульсу, т. е. измеряемой физической величине. В общем же выражении (17) второй член представляет собой компоненту векторного потенциала, являющегося не физической величиной, а функцией, *ротор* которой равен магнитной индукции **B**. Таким образом, электроннооптический показатель преломления представляет собой в общем случае не физическую величину, а функцию Лагранжа *). Однако, как уже упоминалось, он является физической величиной в частном случае чисто электрического поля. В обоих случаях к этим выражениям можно добавить компоненту градиента в направлении движения от произвольной функции координат, не изменяя при этом ни одного физического следствия.

Существует более важное различие между частным и общим случаями, которое проще пояснить, если провести дальнейшие упрощения и свести четырехпараметрическое (∞^4) семейство траекторий к двухпараметрическому (∞^2). Для этого выберем такие траектории, которые начинаются на какой-то поверхности $\mathcal{S}(x, y, z) = \mathcal{S}_0$ и нормальны к ней. В геометрической оптике эту поверхность можно считать волновым фронтом; нетрудно показать, что пучок траекторий остается ортогональным семейству поверхностей $\mathcal{S}(x, y, z) = \text{const}$ (теорема Малюса и Дюпина, см. п. 3.3.3).

Как было показано выше (см. пп. 2 и 3 приложения 1), какая-то «ортогональность» существует и в общем случае, однако ее физический смысл сложнее смысла теоремы Малюса — Дюпина. В этом случае семейству поверхностей $\mathcal{S} = \text{const}$ ортогонален не единичный вектор **s** в направлении движения, а импульс **p**. Поскольку векторный потенциал не определен однозначно, существует бесконечное множество семейств ортогональных поверхностей; однако при наличии магнитного поля никакими калибровочными преобразованиями их нельзя сделать перпендикулярными к траекториям.

На языке геометрии двумерные пучки кривых, ортогональные семейству поверхностей, образуют «нормальную конгруэнцию», в противном случае они

*) Это особенно подчеркивалось в работе [3], где отмечалось также, что если магнитное поле не равно нулю во всем пространстве, то невозможно с помощью калибровочного преобразования так отнормировать потенциал **A**, чтобы он обращался в нуль вместе с магнитным полем.

образуют «косую конгруэнцию» (см. п. 3.2.3). В обычной оптике и электростатической электронной оптике траектории образуют нормальные конгруэнции, а ортогональные им поверхности отождествляются с «волновыми фронтами». При наличии магнитного поля двумерные пучки траекторий образуют косые конгруэнции, к которым нельзя применить понятие ортогональных волновых фронтов. В этом и состоит довольно существенное различие между электронной и обычной оптикой.

3. Волновая механика свободных электронов. В 1905 г. Эйнштейн впервые выдвинул гипотезу о двойственной природе света. Свет распространяется как электромагнитная волна, но при взаимодействии с веществом он ведет себя так, как будто его энергия сконцентрирована в фотонах, каждый из которых несет квант энергии. Вскоре гипотеза Эйнштейна получила блестящее подтверждение при наблюдениях фотоэлектрических и фотохимических процессов.

Предположение о двойственной природе материальных частиц принадлежит Луи де Бройлю, который в 1923 г. показал, что частице с механическим импульсом p_m может соответствовать (релятивистски инвариантно) только длина волны

$$\lambda = h/p_m, \tag{18}$$

где h — универсальная постоянная с размерностью действия, которую де Бройль отождествил с постоянной Планка.

Вскоре после этого Гейзенберг, Борн и Иордан, независимо от де Бройля, предложили первую полную математическую формулировку квантовой механики, однако их методы менее удобны для обсуждения поведения свободных частиц, чем методы волновой механики Шредингера, краткое описание которой и приводится ниже.

Шредингер объединил идеи де Бройля и Гамильтона и пришел к задаче волнового описания движения частиц, которая так же связана с динамикой материальных частиц, как волновая оптика — с геометрической. Его подход, который, как мы теперь знаем, справедлив не всегда, содержал, конечно, целый ряд догадок; в самом деле, хотя геометрическая оптика логически следует из волновой, обратное утверждение неверно.

Предположим, что существует некое волновое поле, интенсивность которого характеризует плотность электронов таким же образом, как интенсивность электромагнитного поля — плотность фотонов. Более того, предположим, что это поле скалярно и его амплитуда описывается скалярной функцией $\Psi(x, y, z, t)$; чтобы учесть волновой характер поля, будем считать, что Ψ удовлетворяет волновому уравнению

$$\nabla^2\Psi = \frac{1}{u^2}\ddot{\Psi},$$

где u — скорость распространения волны, которая в общем случае является функцией координат. Это, конечно, весьма жесткое предположение, так как обычное волновое уравнение с постоянной скоростью распространения можно обобщить множеством различных способов, а данное обобщение является наиболее простым.

Выбирая Ψ в виде «монохроматической волны»

$$\Psi(x, y, z, t) = \psi(x, y, z)\exp(-i\omega t),$$

получим уравнение, не зависящее от времени

$$\nabla^2\psi = -(\omega/u)^2\psi = -(2\pi/\lambda)^2\psi, \tag{19}$$

в которое входит не скорость, а только длина волны λ. Предположим теперь, что λ совпадает с длиной волны де Бройля, если вместо p_m подставить значение импульса, которым бы обладала частица в точке x, y, z согласно законам классической механики. Это значение можно вычислить с помощью (6). Для простоты рассмотрим движение медленного электрона в электростатическом поле.

Тогда из соотношения де Бройля (18) и из формулы (6) имеем

$$\lambda = \frac{h}{\sqrt{2m(E - e\varphi)}}. \tag{20}$$

Подставляя последнее соотношение в (19), получим

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}(E - e\varphi)\psi = 0. \tag{21}$$

Это и есть волновое уравнение Шредингера для свободной частицы в скалярном потенциальном поле. Поскольку оно не зависит от времени, его можно интерпретировать как уравнение, описывающее стационарное (например, периодическое) движение частицы в силовом поле. Однако это уравнение можно использовать и в случае стационарных пучков, с которыми обычно имеют дело в электронной оптике, когда рассматривают много частиц, появляющихся одна за другой, но находящихся в одинаковых условиях. Как в первом, так и во втором случаях разумно предположить в соответствии со статистической интерпретацией Борна, что квадрат модуля $|\psi|^2 = \psi\psi^*$ пропорционален плотности частиц в точке x, y, z, измеренной за длительный промежуток времени, либо, что в данном случае совпадает с этой плотностью, пропорционален вероятности нахождения частицы в данной области пространства в любой момент времени.

Справедливость уравнения (21) была впервые доказана Шредингером, который объяснил с помощью этого уравнения закономерности атомных спектров, предположив, что электрон в атоме находится в таком же силовом поле, что и в старой модели атома Резерфорда — Бора. Для нас больший интерес представляет доказательство справедливости (21) в случае свободных электронов; таким доказательством послужило открытие дифракции электронов Дэвиссоном и Джермером и независимо от них Томсоном в 1927—1928 гг.

Для медленных электронов, обладающих кинетической энергией, эквивалентной eV *эв*, соотношение де Бройля (18) дает следующее приближенное значение длины волны:

$$\lambda = \sqrt{\frac{150}{V}}\ \text{Å}.$$

Таким образом, длины волн электронов, с которыми обычно имеют дело в лабораторных экспериментах, составляют доли ангстрема, т. е. имеют тот же порядок, что и длины волн рентгеновского излучения. Следовательно, волновой характер свободных электронов легче всего продемонстрировать с помощью экспериментов, сходных с опытами по дифракции рентгеновских лучей в кристаллах.

Терминология, применяемая при рентгеновском анализе (которая использовалась также и в случае электронов), несколько отличается от терминологии, принятой в обычной оптике. То, что называется дифракцией рентгеновских лучей или электронов, является на самом деле интерференцией когерентных вторичных волн, испущенных более или менее регулярно расположенными атомами решетки. Дифракция электронов в смысле световой оптики на относительно крупных материальных препятствиях, атомная структура которых не играет никакой роли, происходит на весьма малые углы; впервые ее наблюдал Берш в 1940 г. в электронном микроскопе *) [4].

4. Применение оптических принципов в электроной оптике. Представленный выше элементарный и неполный обзор волновой механики достаточен почти для всех практических нужд электронной оптики. Нет даже необходимости в использовании обобщения волновой механики на случай наличия магнитного поля, поскольку в электронной оптике для исследования дифракционных проблем достаточно приближения Кирхгофа. Допущения, при которых оно

*) Фотографии дифракционных картин Берша воспроизведены в [5].

справедливо, полностью выполняются в электронной оптике. За исключением полей внутри и в непосредственной близости от атомных ядер в электронной оптике не встречаются электрические поля, которые существенно изменяются на расстоянии длины волны используемых электронов. Это утверждение тем более справедливо для магнитных полей, и дифракционные эффекты, обусловленные наличием именно магнитного поля (которых можно ожидать, исходя из волновых уравнений Клейна — Гордона или Дирака), настолько малы, что их невозможно обнаружить в экспериментах со свободными электронами. Таким образом, в электронной оптике можно, правда, с известными оговорками, о которых будет сказано ниже, использовать дифракционную теорию Кирхгофа (см. § 8.3), смысл которой (в несколько обобщенной форме) состоит в следующем:

I. Отмеряя равные оптические пути вдоль лучей, построенных по правилам геометрической оптики, строят волновой фронт от источника до предмета или препятствия.

II. Если препятствием служит светонепроницаемый экран, то часть волнового фронта, не совпадающая с ним, считается невозмущенной как по фазе, так и по интенсивности, а остальная часть — несуществующей. В случае же частично пропускающего излучение препятствия строят по законам геометрической оптики траектории лучей, проходящих через него, причем их изменения по фазе и интенсивности определяются комплексным показателем преломления препятствия. Во всех практических приложениях такой подход оказывается справедливым, поскольку предметы, с которыми приходится иметь дело, например в микроскопии, имеют столь малую толщину, что дифракционными эффектами внутри них можно спокойно пренебречь.

III. Чтобы вычислить дифракционные эффекты в некоторой удаленной точке, например в оптическом изображении, необходимо найти с помощью правил геометрической оптики вклад в эту точку от каждого элемента волнового фронта, выходящего из предмета или препятствия, и просуммировать комплексные амплитуды с учетом их коэффициентов наклона.

Описанный таким образом метод Кирхгофа можно без всяких изменений применить к вопросам электронной оптики, если имеются только электрические поля, в том числе сильные микроскопические поля, обусловленные атомной структурой твердого тела, которые являются в основном электрическими. Однако при наличии магнитных полей уже нельзя строить волновые фронты, отмеряя вдоль траекторий равное число длин волн де Бройля. Как отмечалось выше, волновые фронты при наличии магнитного поля ортогональны не траекториям, а линиям равного полного импульса, или «p-линиям», которые в рамках геометрической оптики не имеют простого истолкования. В принципе волновые фронты можно построить, если начать с какого-нибудь заданного волнового фронта и отмерять оптические пути вдоль «p-линий». Однако такой метод практически почти бесполезен, поскольку он позволяет получить не амплитуды, а только фазы. Эту трудность можно избежать, если мысленно заменить магнитные линзы эквивалентными им электрическими, которые дают такое же изображение, за исключением вращения его как целого и некоторой вращательной дисторсии изображения. Игнорируя подобные специфически магнитные эффекты в процессе распространения и учитывая их только в самом конце, можно с помощью метода Кирхгофа получить результаты, которые справедливы во всех практически встречающихся случаях.

Хотя у электронной и у обычной оптики одинаковое математическое обоснование, их технические средства весьма различны. В основе развития практической световой оптики лежала техника шлифовки и полировки поверхностей необходимой формы из различных прозрачных и отражающих твердых материалов. В электронной же оптике имеется только одна среда — электромагнитное поле. Поэтому в ней всегда возникают трудности при попытках исправ-

ления аберраций, особенно тогда, когда практически невозможно использовать пространственные заряды и токи. В самом начале развития электронной оптики было установлено, что с помощью одних лишь вращательно-симметричных полей нельзя создать рассеивающие линзы и системы, свободные от сферической или хроматической аберраций. Это привело к созданию систем, не обладающих вращательной симметрией, а также систем с искривленными оптическими осями, почти неизвестных в обычной оптике. Следовательно, знание теории оптических инструментов и принципов их конструкций полезно в электронной оптике только при постановке задач; их практические решения будут сильно отличаться от решений соответствующих оптических проблем. Тем не менее очевидно, что «оптическое мышление», основанное на фундаментальной аналогии этих двух областей физики, будет так же полезно в будущем, как и в первом двадцатипятилетии электронной оптики.

ПРИЛОЖЕНИЕ 3

МЕТОД АСИМПТОТИЧЕСКИХ ОЦЕНОК ИНТЕГРАЛОВ

Цель настоящего приложения состоит в изложении математического обоснования до известной степени общих методов, используемых в основной части книги и позволяющих получить асимптотические оценки интегралов некоторых типов, часто встречающихся при решении оптических задач.

1. **Метод наибыстрейшего спуска.** Этот метод*) позволяет получать асимптотические приближения комплексных интегралов типа

$$\int g(z)\exp[kf(z)]\,dz \qquad (1)$$

при больших значениях k, где $g(z)$ и $f(z)$ не зависят от k.

Сейчас и в дальнейшем мы будем считать, что k — вещественная и положительная величина. Полученные результаты, вообще говоря, справедливы и при комплексных значениях k; предварительно скажем несколько слов об асимптотических разложениях функции комплексной переменной, например ζ.

Во-первых, согласно Пуанкаре [7], асимптотическое разложение можно определить следующим образом: если

$$F(\zeta)=\sum_{m=0}^{n}\frac{a_m}{\zeta^m}+R_n(\zeta), \qquad (2)$$

где при всех n величина $\zeta^n R_n(\zeta)$ стремится к 0 при $\zeta\to\infty$ для значений $\arg\zeta$, лежащих в заданном интервале, а $a_0, a_1, \ldots, a_n$ — константы, то

$$F(\zeta)\approx a_0+\frac{a_1}{\zeta}+\frac{a_2}{\zeta^2}+\ldots; \qquad (3)$$

правая часть (3) называется асимптотическим разложением $F(\zeta)$ для заданной области значений $\arg\zeta$.

Если $F(\zeta)$ является частным от деления двух функций, скажем, $G(\zeta)$ и $H(\zeta)$, то

$$G(\zeta)\approx H(\zeta)\left(a_0+\frac{a_1}{\zeta}+\frac{a_2}{\zeta^2}+\ldots\right). \qquad (4)$$

Необходимо подчеркнуть, что в дальнейшем будут рассматриваться только такие разложения, которые имеют вид правой части (4) с $H(\zeta)$, равной $\exp(a\zeta)$, где a — некая постоянная.

Отметим теперь некоторые основные свойства асимптотических разложений **). Если ряд в (3) конечен или сходится при достаточно больших значе-

*) Развитый в основном Дебаем [6].

**) См., например, [8] или [9].

ниях $|\zeta|$, то он является асимптотическим; однако часто подобные ряды сходятся не при всех значениях $|\zeta|$. В общем случае для заданной функции $F(\zeta)$ рассматриваемое разложение справедливо только в каком-то определенном диапазоне значений $\arg\zeta$, если такое разложение возможно при всех значениях $\arg\zeta$, то оно сходится. Кроме того, асимптотическое разложение заданной функции $F(\zeta)$ в соответствующем диапазоне значений $\arg\zeta$ является единственным в том смысле, что коэффициенты в (3) тоже являются единственными, вместе с тем асимптотическое разложение любой функции представляет собой, кроме того, разложение для бесконечного числа функций; например, разложение $F(\zeta)+e^{-\zeta}$ совпадает с асимптотическим разложением $F(\zeta)$ при $-\frac{1}{2}\pi < \arg\zeta < \frac{1}{2}\pi$. Асимптотическое разложение произведения двух функций равно произведению их асимптотических разложений. Наконец, интегрируя (3) почленно, можно получить асимптотическое разложение интеграла от $F(\zeta)$, а дифференцируя почленно — асимптотическое разложение производной от $F(\zeta)$, если она существует.

В общем случае при заданном (достаточно большом) значении $|\zeta|$ абсолютные величины членов разложения (3) сначала спадают до минимума, а затем начинают возрастать. Грубо говоря, если суммировать разложение до какого-нибудь члена, стоящего перед минимальным, то возникающая при этом ошибка окажется порядка первого неучтенного члена *). Очевидно, что чем больше $|\zeta|$, тем больше достигнутая точность. В физических приложениях часто оказывается достаточным учитывать только первый член; например, в теории электромагнетизма поле излучения источника конечных размеров описывается первым членом асимптотического разложения полного поля по отрицательным степеням расстояния до источника.

Основа метода получения асимптотического разложения (1) по отрицательным степеням k заключается в установлении связи подобного разложения с интегралами вида **)

$$\int_0^\infty h(\mu)\exp(-k\mu^2)\,d\mu. \tag{5}$$

Асимптотическое разложение (5) легко получить; для этого функцию $h(\mu)$ нужно разложить в ряд по степеням μ, а затем почленно проинтегрировать его.

В последней фразе заключена сущность *леммы Ватсона* ***) [10], которую можно сформулировать следующим образом. Пусть

$$h(\mu)=\frac{1}{\mu^\alpha}\sum_{s=0}^{\infty} c_s\mu^{\beta s}, \tag{6}$$

причем радиус сходимости этого ряда равен ρ, величина β вещественна и больше нуля, а вещественная часть α меньше единицы. Пусть существует такое вещественное число d, что величина $\mu^\alpha\exp(-d\mu^2)\,h(\mu)$ ограничена при всех вещественных значениях μ, больших ρ. Тогда (буквой Γ обозначается гамма-функция) выражение

$$\frac{1}{2k^{(1-\alpha)/2}}\left\{c_0\Gamma\left(\frac{-\alpha+1}{2}\right)+c_1\Gamma\left(\frac{\beta-\alpha+1}{2}\right)\frac{1}{k^{\beta/2}}+c_2\Gamma\left(\frac{2\beta-\alpha+1}{2}\right)\frac{1}{k^{\beta}}+\right.$$
$$\left.+c_3\Gamma\left(\frac{3\beta-\alpha+1}{2}\right)\frac{1}{k^{3\beta/2}}+\dots\right\} \tag{7}$$

является асимптотическим разложением (5).

---

*) Для ознакомления с методами вычисления погрешностей см. [8].

**) Выбор μ^2 в экспоненте, стоящей под интегралом, вместо первой или любой другой степени μ несуществен, четная степень μ удобна в том отношении, что тогда интеграл типа (1) с пределами от $-\infty$ до ∞ можно преобразовать в интеграл типа (5) с нижним пределом, равным $-\infty$.

***) См. также [8].

Для физических приложений желательно, чтобы $\alpha = 0$ и $\beta = 1$. Как уже отмечалось ранее, первый член в (7) тоже часто дает хорошее приближение.

Для того чтобы с помощью замены переменной интегрирования интеграл (1) можно было представить в виде одного или нескольких интегралов типа (5), необходимо, чтобы путь интегрирования в (1) состоял из отрезков, вдоль которых мнимая часть $f(z)$ остается постоянной, а вещественная монотонно убывает до $-\infty$. Если путь интегрирования не удовлетворяет этому требованию, его необходимо соответствующим образом деформировать. Такая процедура подчиняется, конечно, основным законам интегрирования в комплексной плоскости; здесь будет только показано, каким образом можно замкнуть путь интегрирования с помощью отрезков, обладающих требуемыми свойствами, если предположить, что вычисление любого остающегося контурного интеграла можно провести обычным способом.

Сделаем несколько вполне общих замечаний относительно путей интегрирования, вдоль которых мнимая часть функции $f(z)$ постоянна независимо от вида самой функции. Иными словами, укажем способ соединения двух точек на комплексной плоскости, обладающий отмеченным свойством, хотя этот метод может оказаться непригодным, если $f(z)$ имеет особенности. Важную роль будут играть точки, в которых

$$\frac{df}{dz} = 0. \tag{8}$$

Они называются *седловыми точками*, потому что в них вещественная и мнимая части $f(z)$ стационарны на комплексной плоскости, не будучи ни абсолютными максимумами, ни абсолютными минимумами *).

Если теперь представить $f(z)$ в виде

$$f(z) = u(x, y) + iv(x, y), \tag{9}$$

то легко показать с помощью соотношений Коши — Римана, что вдоль любого пути $v(x, y) = \text{const}$ скорость изменения $u(x, y)$ обращается в нуль лишь в седловой точке. Другими словами, $u(x, y)$ строго монотонна вдоль пути $v(x, y) = \text{const}$, не проходящего через седловую точку, а вдоль пути $v(x, y) = \text{const}$, проходящего через одну или несколько седловых точек, $u(x, y)$ строго монотонна между соседними седловыми точками, а также между крайними седловыми точками и соответствующими концами пути интегрирования. Для любой точки направление, вдоль которого $u(x, y)$ спадает быстрее всего, совпадает с направлением, вдоль которого $v(x, y) = \text{const}$, и в этом смысле такие пути называются путями *наибыстрейшего спуска*. Если функция $f(z)$ однозначна или превращена в однозначную с помощью разреза, то из произвольной точки (x_0, y_0) всегда можно выбрать такой путь интегрирования $v(x, y) = v(x_0, y_0)$, который непрерывен в направлении уменьшения $u(x, y)$ и заканчивается либо на бесконечности, либо в особой точке. Легко показать, что единственными точками (не считая особых точек), где путь $v(x, y) = \text{const}$ может разветвляться, являются точки, в которых $df/dz = 0$; таким образом, если путь $v(x, y) = v(x_0, y_0)$, вдоль которого $u(x, y)$ спадает, был начат правильно, то в дальнейшем можно не беспокоиться о наличии другого пути, пока не встретится седловая точка, из которой имеется по крайней мере одно направление спадания $u(x, y)$.

Предположим теперь, что концевыми точками заданного пути интегрирования в (1) служат A и B и что можно найти пути наибыстрейшего спуска от A и B до бесконечности. Если оба пути оканчиваются на бесконечности в одной и той же области сходимости интеграла, то процедуру можно считать завершенной; однако если они оканчиваются в разных областях сходимости, то

*) Вещественная и мнимая части аналитической функции не имеют абсолютных максимумов и минимумов на комплексной плоскости.

последние нужно соединить или путем $v(x, y) = \mathrm{const}$, вдоль которого скорость изменения $u(x, y)$ меняет знак только один раз (в седловой точке), или в случае необходимости несколькими путями, проходящими через промежуточные области на бесконечности. Асимптотическое разложение в обобщенном смысле (4) можно получить для каждого из этих различных путей интегрирования, однако правильное асимптотическое разложение исходного интеграла соответствует пути, на котором $u(x, y)$ достигает своего наибольшего значения *). Аналогичные замечания справедливы и в том случае, когда пути наибыстрейшего спуска оканчиваются в особых точках.

Необходимо отметить, что в ряде случаев описанный метод не годится, но их исследование становится возможным при небольшом его видоизменении. Прежде всего отметим, что замена верхнего предела в (5) любым положительным числом (не зависящим от k) не приводит к изменению асимптотического разложения (7). Следовательно, этот случай можно исследовать, используя путь наибыстрейшего спуска, вдоль которого $u(x, y)$ стремится не к $-\infty$, а к конечному значению. Как и раньше, можно использовать путь $v(x, y) = \mathrm{const}$, на котором лежит несколько точек, где скорость изменения $u(x, y)$ меняет знак.

Наиболее распространенным случаем, к которому применим метод наибыстрейшего спуска, является тот, где путь интегрирования $v(x, y) = \mathrm{const}$ идет от седловой точки до бесконечности, причем вдоль него $u(x, y)$ все время монотонно спадает **). В этом случае легко получается хорошо известная формула для первого члена асимптотического разложения. Положим, что седловая точка находится в z_0, и произведем в (1) следующую замену переменных:

$$f(z) = f(z_0) - \mu^2. \tag{10}$$

Тогда (1) примет вид

$$-2\exp[kf(z_0)]\int\limits_0^\infty \frac{g(z)}{f'(z)}\mu\exp(-k\mu^2)\,d\mu. \tag{11}$$

Для получения первого члена асимптотического разложения (11) надо знать величину $\mu g(z)/f'(z)$ при $\mu = 0$ $(z = z_0)$. Легко показать, что эта величина равна $g(z_0)/\sqrt{-2f''(z_0)}$, если $g(z_0)$ не стремится к бесконечности, а $f''(z_0) \neq 0$; знак корня выбирается в каждом отдельном случае особо. Тогда искомый первый член разложения имеет вид

$$\sqrt{-\frac{\pi}{2f''(z_0)}}\,g(z_0)\frac{\exp[kf(z_0)]}{\sqrt{k}}. \tag{12}$$

Необходимо сделать несколько замечаний относительно случаев, когда приближение (12) недостаточно или несправедливо. Полное асимптотическое разложение получается, конечно, если разложить $\mu g(z)/f'(z)$ в степенной ряд по μ, а затем почленно проинтегрировать (11). Как видно из (7), степень k в первом члене асимптотического разложения определяется степенью μ, с которой начинается разложение в ряд $\mu g(z)/f'(z)$. Если первый член такого ряда равен $A\mu^{-p}$, где $p < 1$, что обеспечивает сходимость интеграла, то первый

---

*) Возможны, конечно, случаи, когда $u(x, y)$ достигает наибольшего значения на нескольких путях интегрирования.

**) Строго говоря, наиболее распространенным является случай, когда путь интегрирования начинается и оканчивается на бесконечности, а $u(x, y)$ сначала монотонно нарастает от $-\infty$ до максимума в седловой точке, а затем монотонно спадает до $-\infty$. В этом случае выражение (12) необходимо умножить на 2, однако для наших целей удобнее выбрать путь, начинающийся в седловой точке. Поскольку случаи, в которых путь интегрирования начинается в седловой точке или проходит через нее, столь распространены, метод наибыстрейшего спуска иногда называют *методом седловой точки*.

член асимптотического разложения имеет вид

$$-A\Gamma\left(\frac{-p+1}{2}\right)\frac{\exp\,[kf(z_0)]}{k^{(1-p)/2}}. \qquad (13)$$

Таким образом, если $g(z_0)\to\infty$ или $f''(z_0)=0$, то выражением (12) пользоваться нельзя и его следует заменить на (13). В этом случае при $k\to\infty$ множитель перед $\exp\{kf(z_0)\}$ стремится к нулю медленнее, чем $k^{-1/2}$. Если $g(z_0)=0$ или $f''(z_0)\to\infty$, то (12) нельзя считать пригодным, но при $k\to\infty$ множитель перед $\exp\{kf(z_0)\}$ стремится к нулю быстрее, чем $k^{-1/2}$.

Легко показать, что когда путь интегрирования, являющийся путем наибыстрейшего спуска, оканчивается, как и раньше, на бесконечности, но начинается не в седловой точке, то в общем случае неэкспоненциальная часть в первом члене асимптотического разложения пропорциональна не $k^{-1/2}$, как в (12), а k^{-1}. Если же величина $g(z)/f'(z)$ имеет особенность или равна нулю в концевой точке, то, как и ранее, степень k зависит от порядка особенности или нуля.

До сих пор были приведены результаты, получающиеся методом наибыстрейшего спуска, для асимптотических разложений в строгом математическом смысле, причем считалось, что величина k может становиться бесконечно большой, а остальные параметры имеют определенные значения, не зависящие от k. Однако было показано, что вид асимптотического разложения зависит от некоторых условий, в частности от того, начинается ли путь наибыстрейшего спуска в седловой точке или нет. Иными словами, вид разложения может зависеть не только от k, но и от других параметров, и может резко меняться, когда эти параметры принимают определенные критические значения *). Таким образом, для любого данного значения k независимо от того, насколько оно велико, приведенное выше выражение не дает хорошего численного приближения, если значения остальных параметров *достаточно* близки к критическим. Следовательно, практически необходимо получить выражения, обеспечивающие плавный переход от одного асимптотического разложения к другому. Естественно, что эти выражения представляют собой более сложные функции k, чем (13), однако имеет смысл рассмотреть три случая, которые можно исследовать довольно строго. В кратком обзоре, излагаемом ниже, везде неявно предполагается, что комплексность всех величин такая же, как и раньше.

Предположим сначала, что $f''(z_0)$ обращается в нуль. Тогда (12) служит хорошим приближением лишь для случая больших k, и необходимо выражение, с помощью которого можно перейти от (12) к другой форме, пригодной при $f''(z_0)=0$. Поскольку величина $f''(z_0)$ близка к нулю, около z_0 должна существовать вторая седловая точка, скажем, z_1. Тогда искомое выражение получается с помощью преобразования

$$f(z)=\alpha-\beta\mu+\frac{1}{3}\mu^3,$$

где

$$\alpha=\frac{1}{2}\,[f(z_1)+f(z_0)],\qquad \frac{2}{3}\beta^{3/2}=\frac{1}{2}\,[f(z_1)-f(z_0)].$$

Это преобразование представляет z в качестве регулярной функции μ в окрестности z_0 и z_1 и приводит к асимптотическому разложению, выраженному через интеграл Эйри **) и его первую производную по аргументу $k\beta$.

---

*) По сути дела такое поведение обусловлено *неравномерной* сходимостью. Например, при $k\to\infty$ величина $\frac{1}{1+ka}$ стремится либо к $1/ka$ при $a\neq 0$, либо к 1 при $a=0$.

**) Роль интеграла Эйри в однородных аппроксимациях такого типа была обнаружена уже давно (см., например, [8]), но лишь сравнительно недавно был разработан удовлетворительный метод его использования (см. [11]).

Предположим далее, что разложение $\mu g(z)/f'(z)$ в степенной ряд по μ имеет радиус сходимости, стремящийся к нулю, так как $g(z)$ обладает простым полюсом, находящимся вблизи седловой точки. Тогда (12) снова служит лишь хорошим приближением для случая больших k, и требуется найти выражение, с помощью которого можно перейти от (12) к другой форме, пригодной при совпадении полюса с седловой точкой. Последний случай исследовался многими авторами [12—16]. Оказалось, что неэкспоненциальную часть подынтегрального выражения следует представить в виде суммы двух членов, один из которых содержит только полюс, тогда как другой не имеет особенностей. С последним членом обходятся обычным образом, а первый сводится к интегралу Френеля (или интегралу ошибок) в общем случае от комплексного аргумента.

Наконец, предположим, что начальная точка пути наибыстрейшего спуска не совпадает с седловой, но приближается к ней весьма близко. Ясно, что в этом случае для перехода от одной асимптотической формы к другой снова можно воспользоваться интегралом ошибок [14].

2. Метод стационарной фазы. Метод стационарной фазы отличен от метода наибыстрейшего спуска, хотя они и весьма сходны между собой. Метод стационарной фазы менее общий, и его труднее доказать аналитически, но он часто теснее связан с физической задачей. Подлежащий исследованию интеграл удобнее записать в виде

$$\int g(z)\exp[ikf(z)]\,dz, \tag{14}$$

чем в форме (1), поскольку экспонента обычно описывает бегущую волну.

В обозначениях (9) кривая $v(x, y) = \mathrm{const}$ снова описывает путь интегрирования; однако в (14), в противоположность методу наибыстрейшего спуска, амплитудная часть экспоненты остается постоянной вдоль этого пути, тогда как фаза меняется с максимальной скоростью. Можно, как и раньше, показать, что основные вклады в интеграл вносят отрезки пути, лежащие вблизи седловых и концевых точек, однако физическое толкование этого результата проводится теперь не в терминах спадания амплитуды, а в терминах «фазовой интерференции» (см. § 8.3).

Метод стационарной фазы был впервые использован Кельвином [17]. Строгое математическое рассмотрение изложенных выше утверждений принадлежит Ватсону [18]; оно основано на том, что если $0<m<1$, a — положительная постоянная и полная флуктуация функции $F(x)$ ограничена при $x \gg 0$, то *) при $k \to \infty$

$$k^m \int_0^a x^{m-1}F(x)\exp(ikx)\,dx \to F(0)\,\Gamma(m)\exp\left(\frac{1}{2}\,im\pi\right).$$

Исследование Ватсона имеет, однако, весьма ограниченное применение. В частности, с его помощью, по-видимому, невозможно провести полное асимптотическое разложение. Полное разложение было подробно исследовано в работе Фокке [20] для случая, когда $f(z)$ — вещественная функция, а путь интегрирования совпадает с действительной осью. Фокке использовал метод нейтрализующей функции, предложенный ранее Вандер-Корпутом [21].

Следствия, вытекающие из метода стационарной фазы и из метода наибыстрейшего спуска, весьма сходны. Так, если путь интегрирования в (14) начинается в седловой точке z_0 и уходит в бесконечность вдоль кривой $v(x, y) = \mathrm{const}$, не встречая другой седловой точки, то приближение, соответствующее (12), имеет вид

$$\sqrt{-\frac{\pi}{2f''(z_0)}}\,g(z_0)\exp\left(-\frac{1}{4}\,i\pi\right)\frac{\exp(ikf(z_0))}{\sqrt{k}}. \tag{15}$$

*) Этот результат принадлежит Бромвичу [19].

Однако здесь необходимо отметить одно различие между обоими методами. При использовании метода наибыстрейшего спуска с путем интегрирования, начинающимся в седловой точке и не уходящим на бесконечность, вклад в асимптотическое разложение от концевой точки пути оказывается бесконечно малым по сравнению со вкладом от седловой точки, поскольку первый содержит дополнительный экспоненциальный множитель. Вместе с тем при использовании метода стационарной фазы вклад от концевой точки пути интегрирования равен по порядку величины вкладу от седловой точки, деленному на $k^{1/2}$; поэтому он не входит в асимптотическое приближение только в том случае, если учитывается лишь первый член разложения.

Таким образом, методы наибыстрейшего спуска и стационарной фазы (если отвлечься от их математического представления) состоят в выборе такого пути интегрирования, вдоль которого подынтегральное выражение, содержащее экспоненциальный множитель, вносит пренебрежимо малый вклад в интеграл везде, за исключением окрестностей некоторых *критических точек*, являющихся либо седловыми, либо концевыми точками пути интегрирования.

3. Двойные интегралы. В § 8.3 и 9.1 было показано, что решение задачи о дифракции электромагнитного поля на отверстии сводится к вычислению двойных интегралов типа

$$\iint g(x, y) \exp[ikf(x, y)]\,dx\,dy, \tag{16}$$

где $g(x, y)$ и $f(x, y)$ не зависят от k, а область интегрирования определяется отверстием. Ясно, что интеграл (16) сходен с (14), и его приближенное значение при больших k можно также получить с помощью асимптотических разложений.

Естественно, что теория асимптотических разложений двойных интегралов намного сложнее, чем в случае однократных интегралов. Техника интегрирования на комплексной плоскости непосредственно применима лишь для однократных интегралов, поэтому представляется, что для вычисления двойных интегралов следует воспользоваться методами, отличными от изложенных выше.

Случай, когда $f(x, y)$ является вещественной функцией, подробно исследовался Фокке с помощью нейтрализующей функции в работе [20] *); мы уже ссылались на нее в связи с применением этого метода к однократным интегралам. Анализ показывает, что вклады в асимптотическое разложение вносят лишь области в окрестностях определенных *критических точек* и что при различных типах таких точек **) в главных членах соответствующих вкладов появляются разные степени k.

Существуют три типа критических точек. Исследуем кратко главные члены их вкладов в асимптотическое разложение, не учитывая случаев, когда критическая точка принадлежит одновременно к нескольким типам.

Критической точкой первого рода является точка, лежащая в области интегрирования, в которой

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0. \tag{17}$$

Вблизи этой критической точки, скажем в точке (x_0, y_0), имеем

$$f(x, y) = f(x_0, y_0) + \frac{1}{2}\alpha(x-x_0)^2 + \frac{1}{2}\beta(y-y_0)^2 + \gamma(x-x_0)(y-y_0) + \ldots, \tag{18}$$

где $\alpha = \partial^2 f/\partial x^2$, $\beta = \partial^2 f/\partial y^2$, $\gamma = \partial^2 f/\partial x\,\partial y$, причем частные производные берутся в точке (x_0, y_0). Выберем теперь такие новые переменные интегрирования ξ, η,

*) См. также [22, 23, 23а].

**) Критические точки для двойных интегралов упоминаются в работе [21]; их качественное значение для задач оптики отмечено в статье [24].

что

$$f(x,\ y)=f(x_0,\ y_0)+\frac{1}{2}\alpha\xi^2+\frac{1}{2}\beta\eta^2+\gamma\xi\eta. \qquad (19)$$

Тогда искомое асимптотическое приближение интеграла (16) примет вид

$$g(x_0,\ y_0)\exp\left[ikf(x_0,\ y_0)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[\frac{1}{2}ik(\alpha\xi^2+\beta\eta^2+2\gamma\xi\eta)\right]d\xi\,d\eta=$$

$$=\frac{2\pi i\sigma}{\sqrt{|\alpha\beta-\gamma^2|}}g(x_0,\ y_0)\frac{\exp\left[ikf(x_0,\ y_0)\right]}{k}; \qquad (20)$$

квадратный корень *) во второй строчке выражения (20) берется со знаком плюс, и

$$\sigma=\begin{cases}+1 & \text{при}\quad \alpha\beta>\gamma^2,\quad \alpha>0,\\ -1 & \text{при}\quad \alpha\beta>\gamma^2,\quad \alpha<0,\\ -i & \text{при}\quad \alpha\beta<\gamma^2.\end{cases} \qquad (21)$$

Выражение (20) является аналогом асимптотического приближения (15) для однократного интеграла.

Критическими точками второго рода служат точки на кривой, ограничивающей область интегрирования, в которых $\partial f/\partial s=0$, где s есть элемент дуги этой кривой. В отличие от (20), степень k в неэкспоненциальной части главного члена соответствующего вклада в асимптотическое разложение равна не -1, а $-3/2$.

Наконец, *критическими точками третьего рода* служат угловые точки на кривой, ограничивающей область интегрирования, т. е. точки, в которых угол наклона кривой терпит разрыв. В этом случае соответствующий множитель равен k^{-2}.

ПРИЛОЖЕНИЕ 4

ДЕЛЬТА-ФУНКЦИЯ ДИРАКА

В настоящем приложении приводится сводка основных свойств дельта-функции **), которая оказалась полезной при описании точечных источников, точечных зарядов и т. д. Эта функция, используемая особенно часто как в квантовой механике, так и в классической прикладной математике, определяется с помощью следующих уравнений:

$$\delta(x)=0 \quad \text{при} \quad x\neq 0, \qquad (1а)$$

$$\int_{-\infty}^{\infty}\delta(x)\,dx=1. \qquad (1б)$$

Очевидно, что $\delta(x)$ не является функцией в обычном математическом смыс-

*) Член, стоящий под корнем, имеет простое геометрическое истолкование. Рассмотрим поверхность $z=f(x,\ y)$. Пусть R_1 и R_2 — главные радиусы кривизны и $K=1/R_1R_2$ — гауссова кривизна в какой-либо точке нашей поверхности. Тогда (см., например, [25])

$$K=\frac{(f_{xx}f_{yy}-f_{xy}^2)}{(1+f_x^2+f_y^2)^2},$$

где $f_x=\partial f/\partial x$, $f_{xx}=\partial^2 f/\partial x^2$ и т. д. В критической точке первого рода $f_x=f_y=0$, $f_{xx}=\alpha$ и т. д., и полученное выражение принимает вид

$$K=\alpha\beta-\gamma^2.$$

**) Эта функция иногда называется *импульсной функцией*. Она получила широкую известность после опубликования работы Дирака [26], но о ней знали математики и физики значительно раньше, в основном благодаря работам Хевисайда (см. [27]).

ле *), поскольку, если функция равна нулю везде, за исключением одной точки, а интеграл от нее существует, то этот интеграл обязательно должен равняться нулю.

Рассмотрим набор функций $\delta(x, \mu)$, которые при увеличении μ отличаются заметно от нуля лишь на все меньшем и меньшем x-интервале около начала координат, а их интеграл для всех значений μ равен единице; т. е.

$$\int_{-\infty}^{\infty} \delta(x, \mu)\, dx = 1. \qquad (2)$$

Такими функциями служат, например, следующие (рис. 8):

$$\delta(x, \mu) = \frac{\mu}{\sqrt{\pi}} \exp(-\mu^2 x^2). \qquad (3)$$

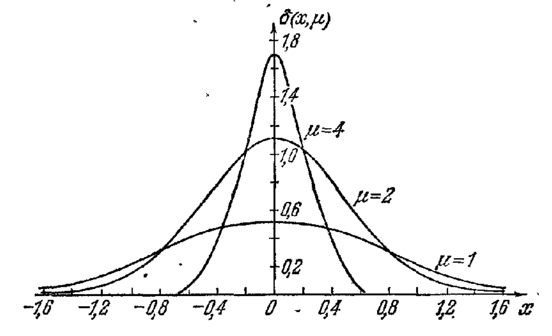


Рис. 8. К определению дельта-функции Дирака.

Графики функции $\delta(x, \mu) = \frac{\mu}{\sqrt{\pi}} \exp(-\mu^2 x^2)$ при $\mu = 1, 2, 4$.

Полная площадь под каждой кривой равна единице.

Соблазнительно попытаться определить дельта-функцию Дирака как предел функций (3) при $\mu \to \infty$; однако следует иметь в виду, что такой предел существует не при всех x. Предел же

$$\lim_{\mu \to \infty} \int_{-\infty}^{\infty} \delta(x, \mu)\, dx \qquad (4)$$

существует всегда и равен единице. *Будем интерпретировать любую операцию над $\delta(x)$ как операцию над какой-то $\delta(x, \mu)$ из соответствующего набора, например из* (3), *с последующим нахождением предела при $\mu \to \infty$ в конце вычислений.* Очевидно, что при такой интерпретации соотношение (1б) выполняется. Явный вид функций $\delta(x, \mu)$ несуществен при условии, что их колебания (если они существуют) около начала координат не слишком резки.

Важным свойством дельта-функции Дирака является следующее:

$$\int_{-\infty}^{+\infty} f(x)\, \delta(x-a)\, dx = f(a). \qquad (5)$$

Здесь $f(x)$ — любая непрерывная функция x. Справедливость (5) сразу же становится очевидной, если заменить $\delta(x-a)$ на $\delta(x-a, \mu)$ и исследовать поведение интеграла при больших значениях μ. Ясно, что при $\mu \to \infty$

$$\int_{-\infty}^{\infty} f(x)\, \delta(x-a, \mu)\, dx \qquad (6)$$

существенно зависит от значений $f(x)$ в окрестности точки $x = a$, и погрешность, возникающую при замене $f(x)$ на $f(a)$, можно сделать пренебрежимо малой, взяв достаточно большое μ. Используя затем (1б), получим (5). Отсюда следует, что умножение непрерывной функции на $\delta(x-a)$ и последующее ее интегрирование по всем значениям x эквивалентно замене ее аргумента на a.

*) Теорию дельта-функции можно сделать математически строгой путем использования понятий распределений (см. книгу Шварца [28]). Более простой вариант теории Шварца был предложен Темплом [29] и описан в книге [30].

Фактически тот же результат получится, если интегрировать функцию не от $-\infty$ до $+\infty$, а по любой области, содержащей точку $x=a$. Полученный результат можно символически записать еще и в виде

$$f(x)\,\delta(x-a)=f(a)\,\delta(x-a); \tag{7}$$

при такой записи мы видим, что правая и левая части (7) дают одинаковый результат после интегрирования. В частности, при $f(x)=x$, $a=0$ соотношение (7) дает

$$x\delta(x)=0. \tag{8}$$

Аналогичным образом легко доказать справедливость следующих выражений:

$$\delta(-x)=\delta(x), \tag{9}$$

$$\delta(ax)=\frac{1}{|a|}\,\delta(x), \tag{10}$$

$$\delta(x^2-a^2)=\frac{1}{2|a|}\,\{\delta(x-a)+\delta(x+a)\}, \tag{11}$$

$$\int\limits_{-\infty}^{+\infty}\delta(a-x)\,\delta(x-b)\,dx=\delta(a-b). \tag{12}$$

Для доказательства справедливости, например, (10) сравним интегралы от $f(x)\,\delta(ax)$ и $f(x)\frac{1}{|a|}\delta(ax)$. Имеем

$$\int\limits_{-\infty}^{+\infty}f(x)\,\delta(ax)\,dx=\pm\int\limits_{-\infty}^{+\infty}f\left(\frac{y}{a}\right)\delta(y)\,\frac{1}{a}\,dy=\frac{1}{|a|}\,f(0),$$

где второй интеграл берется со знаком плюс или минус в зависимости от того, положительно ли a, отрицательно или равно 0. Из (5) следует, что

$$\int\limits_{-\infty}^{+\infty}f(x)\,\frac{1}{|a|}\,\delta(x)\,dx=\frac{1}{|a|}\,f(0).$$

Как мы видим, эти интегралы совпадают, что и означает справедливость (10). Аналогично соотношение (12) означает, что если умножить обе его части на непрерывную функцию от a или b и проинтегрировать соответственно по всем значениям a или b, то мы получим тождество.

Выясним далее, как можно интерпретировать *производные* от дельта-функции. Используя «аппроксимирующие функции» $\delta(x,\mu)$ и интегрируя по частям, получим

$$\int\limits_{-\infty}^{+\infty}f(x)\,\delta'(x,\mu)\,dx=f(\infty)\,\delta(\infty,\mu)-f(-\infty)\,\delta(-\infty,\mu)-\int\limits_{-\infty}^{+\infty}f'(x)\,\delta(x,\mu)\,dx.$$

При переходе к пределу при $\mu\to\infty$ первые два члена в правой части исчезают, и мы имеем

$$\int\limits_{-\infty}^{+\infty}f(x)\,\delta'(x)\,dx=-f'(0). \tag{13}$$

Повторяя этот процесс, найдем

$$\int\limits_{-\infty}^{+\infty}f(x)\,\delta^{(n)}(x)\,dx=(-1)^n f^{(n)}(0). \tag{14}$$

Легко доказать справедливость следующих соотношений:

$$\delta'(-x)=-\delta'(x), \tag{15}$$

$$x\delta'(x)=-\delta(x). \tag{16}$$

Часто оказывается удобным (см., например, приложение 6) выразить дельта-функцию Дирака через *единичную функцию Хевисайда* (называемую иногда *ступенчатой функцией*) $U(x)$, определяемую соотношениями

$$\left. \begin{aligned} U(x) &= 0 \quad \text{при} \quad x < 0, \\ U(x) &= 1 \quad \text{при} \quad x > 0. \end{aligned} \right\} \tag{17}$$

Если, как и раньше, обозначить штрихом производную по x, то, интегрируя по частям (при $x_1 > 0$, $x_2 > 0$), формально получим

$$\int_{-x_1}^{x_2} f(x)\, U'(x)\, dx = [f(x)\, U(x)]_{-x_1}^{x_2} - \int_{-x_1}^{x_2} f'(x)\, U(x)\, dx =$$

$$= f(x_2) - \int_0^{x_2} f'(x)\, dx = f(x_2) - f(x_2) + f(0) = f(0).$$

Полагая $x = y - a$, $f(x) = f(y - a) = F(y)$ и переходя к пределу $x_1 \to \infty$, $x_2 \to \infty$, имеем

$$\int_{-\infty}^{\infty} F(y)\, U'(y-a)\, dy = F(a),$$

т. е. U' обладает свойством (5). В частности, при $F \equiv 1$ и $a = 0$ находим

$$\int_{-\infty}^{+\infty} U'(y)\, dy = 1;$$

следовательно, U' удовлетворяет соотношению (16). Более того, $U'(x) = 0$, когда $x \neq 0$. Таким образом, производную от единичной функции можно идентифицировать с дельта-функцией, т. е.

$$\delta(x) = \frac{d}{dx} U(x). \tag{18}$$

Дельта-функцию можно также ввести с помощью интеграла Фурье

$$f(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk \int_{-\infty}^{+\infty} f(x) \exp[-ik(x-a)]\, dx. \tag{19}$$

Полагая

$$K(x-a, \mu) = \frac{1}{2\pi} \int_{-\mu}^{\mu} \exp[-ik(x-a)]\, dk = \frac{\sin \mu (x-a)}{\pi (x-a)} \tag{20}$$

и изменяя порядок интегрирования, соотношение (19) можно формально представить в виде

$$f(a) = \int_{-\infty}^{\infty} f(x)\, K(x-a)\, dx, \tag{21}$$

где $K(x-a)$ есть предел $K(x-a, \mu)$ при $\mu \to \infty$. Строго говоря, такой предел в обычном смысле не существует *) при $x - a \neq 0$, однако (21) имеет такой же смысл, как и ранее рассмотренные интегралы, т. е.

$$f(a) = \lim_{\mu \to \infty} \int_{-\infty}^{\infty} f(x)\, K(x-a, \mu)\, dx. \tag{22}$$

Таким образом, K обладает свойством (5). Полагая $f(x) = 1$ в (21), находим, что интеграл от $K(x)$, взятый по всем значениям x, равен единице. Следова-

*) Такой предел существует и равен нулю, если его рассматривать в смысле Чезаро (см. [27]).

тельно, мы получили еще одно представление дельта-функции Дирака, а именно

$$\delta(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp(-ikx)\,dk, \tag{23}$$

т. е. $\delta(x)$ можно рассматривать как фурье-образ от единицы. Существует и обратное соотношение, получающееся из (21), если положить $f(x)=\exp(ikx)$ и $a=0$, а именно

$$\int_{-\infty}^{+\infty}\delta(x)\exp(ikx)\,dx=1. \tag{24}$$

До сих пор мы рассматривали лишь одномерное пространство, но все найденные соотношения легко обобщить на случай нескольких измерений. В частности, рассмотрим трехмерное пространство. Тогда функция

$$\delta(x, y, z)=\delta(x)\,\delta(y)\,\delta(z), \tag{25}$$

обозначаемая часто через $\delta(\mathbf{r})$, где $\mathbf{r}$ — вектор с компонентами x, y, z, удовлетворяет, очевидно, соотношениям, аналогичным (1), т. е.

$$\delta(x, y, z)=0 \quad \text{при} \quad x\neq 0,\quad y\neq 0,\quad z\neq 0, \tag{26а}$$

$$\iiint_{-\infty}^{+\infty}\delta(x, y, z)\,dx\,dy\,dz=1. \tag{26б}$$

Свойство (5) можно представить теперь в виде

$$\iiint_{-\infty}^{+\infty}f(x, y, z)\,\delta(x-a, y-b, z-c)\,dx\,dy\,dz=f(a, b, c), \tag{27}$$

и $\delta(x, y, z)$ удовлетворяет обратным преобразованиям Фурье

$$\delta(x, y, z)=\frac{1}{(2\pi)^3}\iiint_{-\infty}^{+\infty}\exp\left[-i(k_x x+k_y y+k_z z)\right]dk_x\,dk_y\,dk_z, \tag{28}$$

$$\iiint_{-\infty}^{+\infty}\delta(x, y, z)\exp\left[i(k_x x+k_y y+k_z z)\right]dx\,dy\,dz=1. \tag{29}$$

ПРИЛОЖЕНИЕ 5

МАТЕМАТИЧЕСКАЯ ЛЕММА, ИСПОЛЬЗУЕМАЯ ПРИ СТРОГОМ ВЫВОДЕ ЗАКОНА ЛОРЕНТЦ — ЛОРЕНЦА

В настоящем приложении мы докажем лемму, используемую в п. 2.4.2, согласно которой

$$\operatorname{rot}\operatorname{rot}\int_{\sigma}^{\Sigma}\mathbf{Q}(\mathbf{r}')\,G(R)\,dV'\longrightarrow\int_{\sigma}^{\Sigma}\operatorname{rot}\operatorname{rot}\mathbf{Q}(\mathbf{r}')\,G(R)\,dV'+\frac{8\pi}{3}\mathbf{Q}(\mathbf{r}) \tag{1}$$

при $a\to 0$. Здесь $\mathbf{Q}(\mathbf{r})$ — произвольная векторная функция координат и $G(R)=\exp(ikR)/R$. Интегралы берутся по объему, ограниченному снаружи поверхностью Σ и внутри поверхностью сферы σ радиуса a, центр которой находится в точке P с радиусом-вектором $\mathbf{r}(x, y, z)$. R обозначает расстояние $|\mathbf{r}-\mathbf{r}'|$, где $\mathbf{r}'(x', y', z')$ — радиус-вектор элемента объема dV'.

Пусть $\mathbf{A}$ — произвольная векторная функция координат. Компоненты $\operatorname{rot}\operatorname{rot}\mathbf{A}$ имеют вид

$$(\operatorname{rot}\operatorname{rot}\mathbf{A})_x=\frac{\partial^2 A_y}{\partial y\,\partial x}+\frac{\partial^2 A_z}{\partial z\,\partial x}-\frac{\partial^2 A_x}{\partial y^2}-\frac{\partial^2 A_x}{\partial z^2}$$

и т. д., так что

$$\left(\operatorname{rot}\operatorname{rot}\int\limits_{\sigma}^{\Sigma}\mathbf{Q}G\,dV'\right)_x=\frac{\partial^2}{\partial y\,\partial x}\int\limits_{\sigma}^{\Sigma}Q_yG\,dV'+\frac{\partial^2}{\partial z\,\partial x}\int\limits_{\sigma}^{\Sigma}Q_zG\,dV'-\left(\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\int\limits_{\sigma}^{\Sigma}Q_xG\,dV'. \tag{2}$$

Тогда для произвольной дифференцируемой скалярной функции $F(\mathbf{r},\mathbf{r}')$ имеем

$$\frac{\partial}{\partial x}\int\limits_{\sigma}^{\Sigma}F\,dV'=\int\limits_{\sigma}^{\Sigma}\frac{\partial F}{\partial x}dV'+\lim_{\delta x\to 0}\frac{1}{\delta x}\left[\int\limits_{\sigma'}^{\Sigma}F\,dV'-\int\limits_{\sigma}^{\Sigma}F\,dV'\right], \tag{3}$$

где σ' — поверхность небольшой сферы радиуса a с центром в точке $T(x+\delta x,\ y,\ z)$. Для определения предела в (3) заметим, что разность двух интегралов характеризует вклады от двух областей, заштрихованных на рис. 9. Элемент объема можно записать в виде

$$\delta V'=-\delta S'\delta x\rho_x,$$

где $\delta S'$ — элемент поверхности и ρ_x — x-компонента единичного радиального вектора $\boldsymbol{\rho}$, выходящего из точки P. Следовательно,

$$\lim_{\delta x\to 0}\frac{1}{\delta x}\left[\int\limits_{\sigma'}^{\Sigma}F\,dV'-\int\limits_{\sigma}^{\Sigma}F\,dV'\right]=-\int\limits_{\sigma}F\rho_x\,dS'. \tag{4}$$

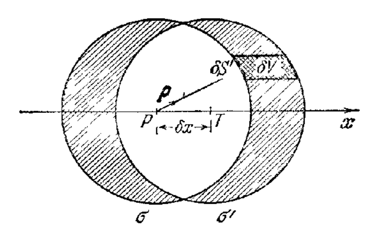


Рис. 9. К вычислению предела $\lim\limits_{\delta x\to 0}\frac{1}{\delta x}\left[\int\limits_{\sigma'}^{\Sigma}F\,dV'-\int\limits_{\sigma}^{\Sigma}F\,dV'\right]$.

Центр σ расположен в точке $P\ (x,\ y,\ z)$, а центр σ' — в точке $T\ (x+\delta x,\ y,\ z)$.

Положив $F=Q_j(\mathbf{r})\,G(R)$, где $Q_j\ (j=x,\ y$ или $z)$ — любая декартова компонента вектора $\mathbf{Q}$, получим из (3) и (4)

$$\frac{\partial}{\partial x}\int\limits_{\sigma}^{\Sigma}(Q_jG)\,dV'=\int\limits_{\sigma}^{\Sigma}\frac{\partial}{\partial x}(Q_jG)\,dV'-\int\limits_{\sigma}(Q_jG)\,\rho_x\,dS'. \tag{5}$$

Рассмотрим далее частные производные второго порядка. Продифференцировав последнее соотношение по x и снова используя его, находим

$$\frac{\partial^2}{\partial x^2}\int\limits_{\sigma}^{\Sigma}(Q_jG)\,dV'=\int\limits_{\sigma}^{\Sigma}\frac{\partial^2}{\partial x^2}(Q_jG)\,dV'-\int\limits_{\sigma}\frac{\partial}{\partial x}(Q_jG)\,\rho_x\,dS'-\frac{\partial}{\partial x}\int\limits_{\sigma}(Q_jG)\,\rho_x\,dS'. \tag{6}$$

Поскольку

$$\left.\begin{aligned}\frac{\partial G}{\partial x}&=\frac{dG}{dR}\frac{\partial R}{\partial x}=-\rho_x\frac{d}{dR}\frac{\exp(ikR)}{R}=\rho_x\left(\frac{1}{R}-ik\right)\frac{\exp(ikR)}{R}\\ \text{и}\qquad dS'&=a^2\,d\Omega,\end{aligned}\right\} \tag{7}$$

где $d\Omega$ — элемент телесного угла, имеем при $a\to 0$

$$\int\limits_{\sigma}\frac{\partial}{\partial x}(Q_jG)\,\rho_x\,dS'=\int\limits_{\sigma}\rho_xQ_j\frac{\partial G}{\partial x}dS'=\int\limits_{\Omega}\rho_x^2Q_j\frac{\exp(ika)}{a}(1-iak)\,d\Omega\to\frac{4\pi}{3}Q_j(\mathbf{r}), \tag{8}$$

здесь Ω — поверхность единичной сферы. Последний интеграл в (6) обращается в нуль при $a\to 0$, в результате чего получим

$$\frac{\partial^2}{\partial x^2}\int\limits_{\sigma}^{\Sigma}(Q_jG)\,dV'\to\int\limits_{\sigma}^{\Sigma}\frac{\partial^2}{\partial x^2}(Q_jG)\,dV'-\frac{4\pi}{3}Q_j(\mathbf{r}). \tag{9}$$

Смешанные вторые частные производные можно вычислить таким же образом. Например, имеем

$$\frac{\partial^2}{\partial y\,\partial x}\int\limits_{\sigma}^{\Sigma}(Q_jG)\,dV' \to \int\limits_{\sigma}^{\Sigma}\frac{\partial^2}{\partial y\,\partial x}(Q_jG)\,dV'. \qquad (10)$$

Член $-\frac{4\pi}{3}Q_j(\mathbf{r})$ теперь отсутствует, поскольку интеграл, соответствующий (8), равен

$$\int\limits_{\sigma}\frac{\partial}{\partial x}(Q_jG)\,\rho_y\,dS' = \int\limits_{\Omega}\rho_x\rho_yQ_j(1-iak)\,d\Omega,$$

т. е. стремится к нулю при $a\to 0$.

Подставляя уравнения вида (9) и (10) в (2), получим при $a\to 0$

$$\left[\operatorname{rot}\operatorname{rot}\int\limits_{\sigma}^{\Sigma}(\mathbf{Q}G)\,dV'\right]_x \to \left[\int\limits_{\sigma}^{\Sigma}\operatorname{rot}\operatorname{rot}(\mathbf{Q}G)\,dV'\right]_x + \frac{8\pi}{3}Q_x. \qquad (11)$$

Аналогичные выражения находим и для y- и z-компонент. Комбинируя эти три соотношения, получим векторную формулу (1).

ПРИЛОЖЕНИЕ 6

РАСПРОСТРАНЕНИЕ РАЗРЫВОВ ЭЛЕКТРОМАГНИТНОГО ПОЛЯ

В п. 3.1.1 отмечалось, что уравнение эйконала в геометрической оптике идентично уравнению, строго описывающему распространение разрывов электромагнитного поля. В более общем виде можно показать, что четыре уравнения (3.1.11а) — (3.1.14а), характеризующие поведение электромагнитного поля, создаваемого геометрическими световыми лучами, совпадают с уравнениями, связывающими векторы поля на движущейся поверхности разрыва. Цель настоящего приложения состоит в математическом доказательстве этой эквивалентности.

1. Соотношения, связывающие разрывы непрерывности векторов поля. В п. 1.1.3 мы рассматривали разрывы векторов поля, возникающие из-за резкого изменения материальных параметров ε и μ, скажем, на поверхности линзы. Разрывы в полях могут также возникать и по совершенно иным причинам, например вследствие того, что источник внезапно начнет излучать. Тогда поле распространяется в пространство, окружающее источник, и с течением времени заполняет все большую и большую область. На границе этой области поле терпит разрыв, причем внутри области векторы поля в общем случае конечны, а вне ее они равны нулю. Установим сначала некоторые общие соотношения, справедливые на любой поверхности, на которой поле терпит разрыв. Предположим для простоты, что для любого момента времени $t>0$ существует только одна такая поверхность; обобщение на случай нескольких поверхностей разрыва (возникающих, например, при отражениях от препятствий, находящихся в среде) не представляет труда.

Пусть $F(x, y, z, t)$ — произвольная поверхность, на которой по крайней мере один из векторов поля терпит разрыв. Если она фиксирована в пространстве, то F, конечно, не зависит от t. Точки по разные стороны от этой поверхности характеризуются неравенствами $F<0$ и $F>0$. Пусть $\mathbf{E}$ — вектор электрического поля, и пусть

$$\left.\begin{aligned}\mathbf{E}(x, y, z, t) &= \mathbf{E}^{(1)}(x, y, z, t), \quad \text{когда} \quad F(x, y, z, t) < 0,\\ \mathbf{E}(x, y, z, t) &= \mathbf{E}^{(2)}(x, y, z, t), \quad \text{когда} \quad F(x, y, z, t) > 0.\end{aligned}\right\} \qquad (1)$$

Тогда $\mathbf{E}$ можно записать в виде

$$\mathbf{E} = E^{(1)}U(-F) + \mathbf{E}^{(2)}U(F), \qquad (2)$$

где U — единичная функция Хевисайда (см. уравнение (17) приложения 4),

С помощью (2) получим выражения для rot **E**, div **E**, $\partial \mathbf{E}/\partial t$ и т. д., входящих в уравнения Максвелла. При дифференцировании суммы или произведения, содержащих разрывный множитель, применим обычные правила дифференцирования и соотношение (18) приложения 4

$$\frac{d}{dx} U(x) = \delta(x), \tag{3}$$

где δ — дельта-функция Дирака. Так, например, из (2) получим

$$\operatorname{rot} \mathbf{E} = U(-F) \operatorname{rot} \mathbf{E}^{(1)} + U(F) \operatorname{rot} \mathbf{E}^{(2)} + \\ + [\operatorname{grad} U(-F)] \times \mathbf{E}^{(1)} + [\operatorname{grad} U(F)] \times \mathbf{E}^{(2)}. \tag{4}$$

Далее имеем

$$\operatorname{grad} U(-F) = -\operatorname{grad} U(F) = -\frac{dU(F)}{dF} \operatorname{grad} F = -\delta(F) \operatorname{grad} F, \tag{5}$$

а из последних двух соотношений следует, что

$$\operatorname{rot} \mathbf{E} = U(-F) \operatorname{rot} \mathbf{E}^{(1)} + U(F) \operatorname{rot} \mathbf{E}^{(2)} + \delta(F) \operatorname{grad} F \times (\Delta \mathbf{E}), \tag{6}$$

где

$$\Delta \mathbf{E} = \mathbf{E}^{(2)} - \mathbf{E}^{(1)}. \tag{7}$$

Аналогичным образом находим

$$\operatorname{div} \mathbf{E} = U(-F) \operatorname{div} \mathbf{E}^{(1)} + U(F) \operatorname{div} \mathbf{E}^{(2)} + \delta(F) \operatorname{grad} F \cdot \Delta \mathbf{E}, \tag{8}$$

$$\frac{\partial \mathbf{E}}{\partial t} = U(-F) \frac{\partial \mathbf{E}^{(1)}}{\partial t} + U(F) \frac{\partial \mathbf{E}^{(2)}}{\partial t} + \delta(F) \frac{\partial F}{\partial t} \Delta \mathbf{E}. \tag{9}$$

Токи и заряды можно представить в таком же виде, как и векторы поля, однако если материальные параметры ε и μ терпят разрыв на поверхности $F = 0$, то в этих выражениях могут появиться дополнительные члены, содержащие поверхностную плотность тока $\hat{\mathbf{j}}$ и поверхностную плотность заряда $\hat{\rho}$. Вклады от этих величин описываются уравнениями (1.1.17а) и (1.1.18а), так что в результате мы получим

$$\mathbf{j} = \mathbf{j}^{(1)} U(-F) + \mathbf{j}^{(2)} U(F) + \hat{\mathbf{j}}\, |\operatorname{grad} F|\, \delta(F), \tag{10}$$

$$\rho = \rho^{(1)} U(-F) + \rho^{(2)} U(F) + \hat{\rho}\, |\operatorname{grad} F|\, \delta(F). \tag{11}$$

Подставим теперь соотношения (6), (8), (9), (10), (11) и аналогичные соотношения для других векторов поля в уравнения Максвелла (1.1.1) — (1.1.4). Члены с индексами (1) сокращаются так же, как и члены с индексом (2), поскольку по обе стороны от поверхности разрыва поля удовлетворяют уравнениям Максвелла. Оставшиеся члены дают следующие *соотношения, связывающие разрывы векторов поля*:

$$\operatorname{grad} F \times \Delta \mathbf{H} - \frac{1}{c} \frac{\partial F}{\partial t} \Delta \mathbf{D} = \frac{4\pi}{c} \hat{\mathbf{j}}\, |\operatorname{grad} F|, \tag{12}$$

$$\operatorname{grad} F \times \Delta \mathbf{E} + \frac{1}{c} \frac{\partial F}{\partial t} \Delta \mathbf{B} = 0, \tag{13}$$

$$\operatorname{grad} F \cdot \Delta \mathbf{D} = 4\pi \hat{\rho}\, |\operatorname{grad} F|, \tag{14}$$

$$\operatorname{grad} F \cdot \Delta \mathbf{B} = 0, \tag{15}$$

Интересно отметить, что эти уравнения получаются формально из уравнений Максвелла, если заменить векторы поля **E**, **H**, **D**, **B** на разности $\Delta\mathbf{E}$, $\Delta\mathbf{H}$, $\Delta\mathbf{D}$, $\Delta\mathbf{B}$, величины $\mathbf{j}$ и ρ на $\hat{\mathbf{j}}$ и $\hat{\rho}$ и дифференциальные операторы $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, $\frac{\partial}{\partial z}$, $\frac{\partial}{\partial t}$ на операторы

$$\frac{1}{|\operatorname{grad} F|} \frac{\partial F}{\partial x}, \quad \frac{1}{|\operatorname{grad} F|} \frac{\partial F}{\partial y}, \quad \frac{1}{|\operatorname{grad} F|} \frac{\partial F}{\partial z}, \quad \frac{1}{|\operatorname{grad} F|} \frac{\partial F}{\partial t}.$$

Пусть $\mathbf{n}_{12}$ — единичный вектор, нормальный к поверхности разрыва и направленный из области $F < 0$ (индекс 1) в область $F > 0$ (индекс 2):

$$\mathbf{n}_{12} = \frac{\operatorname{grad} F}{|\operatorname{grad} F|}. \tag{16}$$

Введем также скорость v, с которой движется поверхность разрыва. Небольшому смещению $\delta\mathbf{r}(\delta x, \delta y, \delta z)$ из точки на поверхности разрыва $F(x, y, z, t)=0$ до точки на соседней поверхности разрыва соответствует такое изменение времени δt, что

$$\operatorname{grad} F\cdot\delta\mathbf{r}+\frac{\partial F}{\partial t}\delta t=0. \tag{17}$$

В частности, для смещения вдоль нормали $\delta\mathbf{r}=\delta s\,\mathbf{n}_{12}$, так что скорость v дается выражением

$$v=\frac{ds}{dt}=-\frac{1}{|\operatorname{grad} F|}\frac{\partial F}{\partial t}. \tag{18}$$

Соотношения (12) — (15) можно записать тогда в виде *):

$$\mathbf{n}_{12}\times\Delta\mathbf{H}+\frac{v}{c}\Delta\mathbf{D}=\frac{4\pi}{c}\hat{\mathbf{j}}, \tag{12a}$$

$$\mathbf{n}_{12}\times\Delta\mathbf{E}-\frac{v}{c}\Delta\mathbf{B}=0, \tag{13a}$$

$$\mathbf{n}_{12}\cdot\Delta\mathbf{D}=4\pi\hat{\rho}, \tag{14a}$$

$$\mathbf{n}_{12}\cdot\Delta\mathbf{B}=0. \tag{15a}$$

Если разрывы векторов полей возникают из-за резкого изменения величин материальных параметров ε и μ на поверхности $F(x, y, z)=0$, положение которой фиксировано в пространстве, то $v=0$ и соотношения (12a) — (15a) сводятся к уравнениям (25), (23), (19) и (15), приведенным в § 1.1.

2. **Поле на движущейся поверхности разрыва.** Рассмотрим движущуюся поверхность разрыва, возникающую из-за присутствия источника, внезапно начинающего излучать. Представим эту поверхность в виде

$$F(x, y, z, t)\equiv\mathscr{S}(x, y, z)-ct=0, \tag{19}$$

где c — скорость света в вакууме. Векторы поля на поверхности разрыва мы будем обозначать маленькими буквами, т. е.

$$\mathbf{e}(x, y, z)=\mathbf{E}\left[x, y, z, \frac{1}{c}\mathscr{S}(x, y, z)\right]; \tag{20}$$

аналогичные выражения справедливы и для других векторов поля. Далее, в области за движущейся поверхностью (скажем, при $F(x, y, z, t)>0$) поле равно нулю, так что, согласно (1) и (7), имеем $\Delta\mathbf{E}=-\mathbf{E}^{(1)}=-\mathbf{e}$ и т. д. Из материальных уравнений (1.1.10) и (1.1.11) следует, что $\mathbf{d}=\varepsilon\mathbf{e}$, $\mathbf{b}=\mu\mathbf{h}$, и если положить $\hat{\mathbf{j}}=0$, $\hat{\rho}=0$, то из (12) — (15) получим

$$\operatorname{grad}\mathscr{S}\times\mathbf{h}+\varepsilon\mathbf{e}=0, \tag{21}$$

$$\operatorname{grad}\mathscr{S}\times\mathbf{e}-\mu\mathbf{h}=0, \tag{22}$$

$$\operatorname{grad}\mathscr{S}\cdot\mathbf{e}=0, \tag{23}$$

$$\operatorname{grad}\mathscr{S}\cdot\mathbf{h}=0. \tag{24}$$

Эти уравнения формально совпадают с основными уравнениями (3.1.11a) — (3.1.14a) геометрической оптики. Следовательно, *векторы поля на движущейся поверхности разрыва подчиняются точно таким же уравнениям, как и гармонические по времени векторы поля в приближении геометрической оптики, причем движущейся поверхности разрыва соответствуют геометрические волновые фронты* **).

---

*) Эти уравнения служат обобщением системы соотношений для разрывов полей, известных, по-видимому, еще Хевисайду. См. также [31—36] и, кроме того, [23a] стр. 37—51 и [33]. Используемый здесь метод вывода этих уравнений принадлежит Бреммеру [37].

**) Можно также показать, что $\mathbf{e}$ и $\mathbf{h}$ удовлетворяют тем же уравнениям переноса (см. (3.1.41), (3.1.42)), что и комплексные векторные амплитуды полей в геометрической оптике. Этот результат был впервые установлен Лунебергом [34]. См. также [23a] стр. 162 и [38].

Очевидно, что движущаяся поверхность разрыва должна удовлетворять уравнению эйконала

$$\left(\frac{\partial \mathcal{S}}{\partial x}\right)^2+\left(\frac{\partial \mathcal{S}}{\partial y}\right)^2+\left(\frac{\partial \mathcal{S}}{\partial z}\right)^2=n^2, \qquad (25)$$

где $n^2=\varepsilon\mu$. Это уравнение выводится так же, как и ранее. Оно служит условием совместности уравнений (21) и (22) и следует из них после исключения **h** или **e** и использования (23) или (24). Согласно (18), (19) и (25) поверхность разрыва движется со скоростью $v=c/n$.

ПРИЛОЖЕНИЕ

КРУГОВЫЕ ПОЛИНОМЫ ЦЕРНИКЕ

В настоящем приложении будут более подробно рассмотрены круговые полиномы, о которых мы кратко говорили в п. 9.2.1. Эти полиномы были впервые введены и исследованы Цернике [39] в его важной работе, посвященной исследованию метода темного поля и фазового контраста; затем они изучались им и Бринкманом [40], а также Нижбером [41]. Эти полиномы были позднее выведены только из требования ортогональности и инвариантности [42]; в нашем изложении мы будем придерживаться в основном последнего исследования.

1. Некоторые общие замечания. Нетрудно показать, что существует бесчисленное множество полных систем полиномов от двух вещественных переменных x и y, ортогональных внутри единичного круга, т. е. удовлетворяющих условию ортогональности

$$\iint\limits_{x^2+y^2\leqslant 1} V^*_{(\alpha)}(x, y)\, V_{(\beta)}(x, y)\, dx\, dy = A_{\alpha\beta}\delta_{\alpha\beta}. \qquad (1)$$

Здесь $V_{(\alpha)}$ и $V_{(\beta)}$ — два произвольных полинома системы, звездочка — комплексное сопряжение, δ — символ Кронекера и $A_{\alpha\beta}$ — нормировочная постоянная, которая будет определена позднее. Круговые полиномы Цернике отличаются от полиномов других систем некоторыми простыми свойствами инвариантности, которые проще всего объяснить в рамках теории групп. Однако с помощью своего рода нормировки можно избежать введения абстрактного формализма теории групп. Рассмотрим сначала такие системы полиномов, которые «инвариантны по форме» относительно поворота координатных осей вокруг начала координат. Такая инвариантность означает, что при любом повороте

$$x'=x\cos\varphi+y\sin\varphi, \qquad y'=-x\sin\varphi+y\cos\varphi \qquad (2)$$

каждый полином $V(x, y)$ переходит в полином такого же вида, т. е. при использовании преобразования (2) V удовлетворяет следующему соотношению:

$$V(x, y)=G(\varphi)\,V(x', y'), \qquad (3)$$

где $G(\varphi)$ — непрерывная периодическая функция угла поворота φ с периодом 2π и $G(0)=1$.

Далее, осуществление двух последовательных поворотов на углы φ_1 и φ_2 эквивалентно одному повороту на угол $\varphi_1+\varphi_2$. Следовательно, из (3) следует, что величина G должна удовлетворять функциональному уравнению

$$G(\varphi_1)\,G(\varphi_2)=G(\varphi_1+\varphi_2). \qquad (4)$$

Общее решение этого уравнения, имеющее период 2π, хорошо известно *):

$$G(\varphi)=\exp(il\varphi). \qquad (5)$$

*) См., например, [43].

Здесь l — любое целое число, положительное, отрицательное или нуль. Подставляя (5) в (3), полагая $x'=\rho$, $y'=0$ и используя (2), получим, что V должен иметь вид

$$V(\rho\cos\varphi,\ \rho\sin\varphi)=R(\rho)\exp(il\varphi), \tag{6}$$

где $R(\rho)=V(\rho,\ 0)$ зависит только от ρ. Разложим теперь $\exp(il\varphi)$ в ряд по степеням $\cos\varphi$ и $\sin\varphi$. Предположим, что V — полином степени n от переменных $x=\rho\cos\varphi$ и $y=\rho\sin\varphi$; тогда из (6) следует, что $R(\rho)$ есть полином по ρ степени n и не содержит степеней ρ, меньших $|l|$. Более того, $R(\rho)$ является, очевидно, четным или нечетным полиномом в зависимости от четности числа l. Система *круговых полиномов Цернике* отличается от всех других подобных систем тем, что она *содержит полином для каждой пары возможных значений n (степень)* и *l (угловая зависимость)*, т. е. для всех целых значений n и l, для которых $n\geqslant 0$, $l\gtreqless 0$, $n\geqslant |l|$, а $n-|l|$ — четное число. Обозначим произвольный полином этой системы следующим образом:

$$V_n^l(\rho\cos\varphi,\ \rho\sin\varphi)=R_n^l(\rho)\exp(il\varphi). \tag{7}$$

Из (1) и (7) следует, что *радиальные полиномы* $R_n^l(\rho)$ удовлетворяют соотношению

$$\int_0^1 R_n^l(\rho)\,R_{n'}^l(\rho)\,\rho\,d\rho=a_n^l\delta_{nn'}, \tag{8}$$

где

$$a_n^l=\frac{A_n^l}{2\pi}. \tag{9}$$

Для любого заданного значения l нижний индекс n может принимать только значения $|l|$, $|l|+2$, $|l|+4$, ... Соответствующая последовательность $R_{|l|}^l(\rho)$, $R_{|l|+2}^l(\rho)$, $R_{|l|+4}^l(\rho)$ ... получается путем ортогонализации степеней

$$\rho^{|l|},\ \rho^{|l|+2},\ \rho^{|l|+4},\ \ldots \tag{10}$$

с весовым множителем ρ на интервале $0\leqslant\rho\leqslant 1$. Далее, поскольку в (10) входит только модуль l, то

$$R_n^{-l}(\rho)=\beta_n^l R_n^l(\rho), \tag{11}$$

где β_n^l — постоянная, зависящая только от нормировки полиномов R_n^{-l} и R_n^l. В частности, она может быть выбрана так, чтобы $\beta_n^l=1$ для всех значений l и n, и тогда

$$V_n^{\pm m}(\rho\cos\varphi,\ \rho\sin\varphi)=R_n^m(\rho)\exp(\pm im\varphi), \tag{12}$$

где $m=|l|$ — неотрицательное целое число.

Система круговых полиномов содержит $\frac{1}{2}(n+1)(n+2)$ линейно независимых полиномов степени $\leqslant n$. Следовательно, любой одночлен x^iy^j ($i\geqslant 0$, $j\geqslant 0$ — целые числа) и любой полином по x и y можно представить в виде линейной комбинации конечного числа круговых полиномов V_n^l. Тогда в соответствии с теоремой Вейерштрасса *) такая система будет *полной*.

2. **Точные выражения для радиальных полиномов $R_n^{\pm m}(\rho)$.** Поскольку $R_n^{\pm m}(\rho)$ есть полином по ρ степени n и он не содержит степени ρ, меньшей m, и является четным или нечетным в зависимости от того, четно или нечетно n, то его можно представить в виде

$$R_n^{\pm m}(\rho)=t^{m/2}Q_{\frac{n-m}{2}}(t), \tag{13}$$

где $t=\rho^2$ и $Q_{\frac{n-m}{2}}(t)$ — полином по t степени $\frac{1}{2}(n-m)$. Согласно (8)

*) См., например, [44].

полиномы Q должны удовлетворять соотношениям

$$\frac{1}{2}\int_0^1 t^m Q_k(t)\,Q_{k'}(t)\,dt = a_n^{\pm m}\delta_{kk'}, \quad \text{где} \quad k=\frac{1}{2}(n-m), \quad k'=\frac{1}{2}(n'-m). \tag{14}$$

Отсюда следует, что полиномы $Q_0(t)$, $Q_1(t)$, ..., $Q_k(t)$, ... можно получить ортогонализацией последовательности натуральных степеней

$$1, \; t, \; t^2, \; \ldots, \; t^k, \; \ldots \tag{15}$$

с весовым множителем $w(t)=t^m$ в области $0 \leqslant t \leqslant 1$. Хорошо известные полиномы Якоби (или гипергеометрические функции) (см. [44])

$$G_k(p,\, q,\, t) = \frac{(q-1)!}{(q+k-1)!}\, t^{1-q}(1-t)^{q-p}\frac{d^k}{dt^k}\left[t^{q-1+k}(1-t)^{p-q+k}\right] = \tag{16}$$

$$= \frac{k!\,(q-1)!}{(p+k-1)!}\sum_{s=0}^{k}(-1)^s\frac{(p+k+s-1)!}{(k-s)!\,s!\,(q+s-1)!}\,t^s \tag{17}$$

($k \geqslant 0$, $q>0$, $p-q>-1$) можно определить как функции, получающиеся при ортогонализации (15) с весовой функцией более общего вида

$$w(t) = t^{q-1}(1-t)^{p-q}$$

в области $0 \leqslant t \leqslant 1$. Свойства ортонормируемости этих полиномов выражаются следующим образом [45]:

$$\int_0^1 t^{q-1}(1-t)^{p-q}G_k(p,\, q,\, t)\,G_{k'}(p,\, q,\, t)\,dt = b_k(p,\, q)\,\delta_{kk'}, \tag{18}$$

где

$$b_k(p,\, q) = \frac{k!\,[(q-1)!]^2\,[p-q+k]!}{[q-1+k]!\,[p-1+k]!\,[p+2k]}. \tag{19}$$

(При таком выборе b_k имеем $G_k(p,\, q,\, 0)=1$ для всех k.) Сравнивая (18) и (14), получим, что*)

$$Q_k(t) = \sqrt{\frac{2a_n^{\pm m}}{b_k(m+1,\; m+1)}}\,G_k(m+1,\; m+1,\; t). \tag{20}$$

Из (13) и (20) вытекает следующее соотношение между радиальными полиномами и полиномами Якоби:

$$R_n^{\pm m}(\rho) = \sqrt{\frac{2a_n^{\pm m}}{b_k[(m+1),\;(m+1)]}}\cdot\rho^m G_k(m+1,\; m+1,\; \rho^2), \quad \left[k=\frac{1}{2}(n-m)\right]. \tag{21}$$

Следуя Цернике, выберем нормировку таким образом, чтобы при всех n и m оставалось справедливым соотношение

$$R_n^{\pm m}(1) = 1. \tag{22}$$

Тогда из (21) и (22) находим

$$\sqrt{\frac{b_k(m+1,\; m+1)}{2a_n^{\pm m}}} = G_k(m+1,\; m+1,\; 1). \tag{23}$$

Значение $G_k(m+1,\; m+1,\; 1)$ можно получить с помощью производящей функции полиномов Якоби [44]. Имеем

$$\frac{\left[z-1+\sqrt{1-2z(1-2\rho^2)+z^2}\right]^m}{(2z\rho^2)^m\sqrt{1-2z(1-2\rho^2)+z^2}} = \sum_{s=0}^{\infty}\binom{m+s}{s}G_s(m+1,\; m+1,\; \rho^2)z^s. \tag{24}$$

*) Знак квадратного корня в правой части (20) определяется из уравнения (26), следующего далее.

При $\rho=1$ левая часть последнего соотношения равна $(1+z)^{-1}$, и разлагая ее в степенной ряд и сравнивая с правой частью, получим

$$G_s(m+1,\ m+1,\ 1)=\frac{(-1)^s}{\binom{m+s}{s}}. \tag{25}$$

Из (25) и (23) следует, что

$$\sqrt{\frac{2a_n^{\pm m}}{b_k(m+1,\ m+1)}}=(-1)^{\frac{n-m}{2}}\binom{{}^1/_2(n+m)}{{}^1/_2(n-m)}; \tag{26}$$

используя (16), (17) и (26), найдем из (21) следующие окончательные выражения для радиальных полиномов:

$$R_n^{\pm m}(\rho)=\frac{1}{\left(\frac{n-m}{2}\right)!\,\rho^m}\left\{\frac{d}{d(\rho^2)}\right\}^{\frac{n-m}{2}}\left\{(\rho^2)^{\frac{m+n}{2}}(\rho^2-1)^{\frac{n-m}{2}}\right\} \tag{27}$$

или

$$R_n^{\pm m}(\rho)=\sum_{s=0}^{\frac{1}{2}(n-m)}(-1)^s\frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\rho^{n-2s}. \tag{28}$$

В табл. 9.1 приведены в явном виде выражения для нескольких первых полиномов.

Для нормировочной постоянной $a_n^{\pm m}$ получим из (26) и (19)

$$a_n^{\pm m}=\frac{1}{2n+2}. \tag{29}$$

Для нахождения производящей функции радиальных полиномов напишем в (21) и (26) s вместо $k=(n-m)/2$ и $m+2s$ вместо n и подставим полученные выражения в (24). Тогда имеем

$$\frac{\left[1+z-\sqrt{1+2z(1-2\rho^2)+z^2}\right]^m}{(2z\rho)^m\sqrt{1+2z(1-2\rho^2)+z^2}}=\sum_{s=0}^{\infty}z^sR_{m+2s}^{\pm m}(\rho). \tag{30}$$

Наконец, вычислим интеграл

$$\int_0^1 R_n^m(\rho)\,J_m(v\rho)\,\rho\,d\rho,$$

который, как мы видели в гл. 9, играет важную роль в дифракционной теории аберраций Цернике—Нижбера. Подставим в этот интеграл выражение (27) для $R_n^m(\rho)$, а вместо функции Бесселя J_m—ее разложение в ряд [44]. В результате получим

$$\int_0^1 R_n^m(\rho)\,J_m(v\rho)\,\rho\,d\rho=$$
$$=\frac{1}{2\left(\frac{n-m}{2}\right)!}\sum_{s=0}^{\infty}\frac{(-1)^s}{s!\,(s+m)!}\left(\frac{v}{2}\right)^{m+2s}f\left(s,\ \frac{n-m}{2},\ \frac{n+m}{2},\ \frac{n-m}{2}\right), \tag{31}$$

где

$$f(s,\ p,\ q,\ r)=\int_0^1 u^s\left(\frac{d}{du}\right)^p\{u^q(u-1)^r\}\,du, \tag{32}$$

а $p,\ q,\ r,\ s$—неотрицательные целые числа. Интегрируя (32) по частям, имеем

$$f(s,\ p,\ q,\ r)=\left\{u^s\left(\frac{d}{du}\right)^{p-1}[u^q(u-1)^r]\right\}_0^1-s\int_0^1 u^{s-1}\left(\frac{d}{du}\right)^{p-1}\{u^q(u-1)^r\}\,du. \tag{33}$$

Если $r \geqslant p$ и $s+q+p \geqslant 0$, то первый член в правой части исчезает, так что

$$f(s, p, q, r) = -sf(s-1, p-1, q, r). \quad (34)$$

Рассмотрим по отдельности случаи $s \geqslant p$ и $s < p$.

Когда $s \geqslant p$, то, применяя (34) p раз, получим

$$f(s, p, q, r) = (-1)^p s(s-1)(s-2)\ldots(s-p+1) f(s-p, 0, q, r) =$$
$$= \frac{(-1)^{p+r} s!}{(s-p)!} \int_0^1 u^{s+q-p}(1-u)^r \, du. \quad (35)$$

Интеграл в правой части есть интеграл Эйлера первого рода (бета-функция), который равен $(s+q-p)!\, r!/(s+q+r-p+1)!$ (см. [44]). Следовательно, при $s \geqslant p$ имеем

$$f(s, p, q, r) = (-1)^{p+r} \frac{s!\,(s+q-p)!\, r!}{(s-p)!\,(s+q+r-p+1)!}. \quad (36)$$

Рассмотрим теперь случай $s < p$. Применяя (34) s раз, получим

$$f(s, p, q, r) = (-1)^s s(s-1)\ldots f(0, p-s, q, r) =$$
$$= (-1)^s s! \left\{ \left(\frac{d}{du}\right)^{p-s-1} [u^q (u-1)^r] \right\}_0^1 = 0. \quad (37)$$

Подставим теперь (36) и (37) в (31) и введем такую новую переменную l, что $s = \frac{1}{2}(n-m)+l$. Тогда

$$\int_0^1 R_n^m(\rho) J_m(v\rho) \rho \, d\rho = \frac{(-1)^{3(n-m)/2}}{v} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!\,(n+l+1)!} \left(\frac{v}{2}\right)^{n+2l+1}. \quad (38)$$

Ряд в правой части (38) есть разложение функции $J_{n+1}(v)$. Поскольку $n-m$ — четное число, то множитель $(-1)^{3(n-m)/2}$ можно заменить на $(-1)^{(n-m)/2}$, в результате чего получим окончательно

$$\int_0^1 R_n^m(\rho) J_m(v\rho) \rho \, d\rho = (-1)^{(n-m)/2} \frac{J_{n+1}(v)}{v}. \quad (39)$$

ПРИЛОЖЕНИЕ 8

ДОКАЗАТЕЛЬСТВО ОДНОГО НЕРАВЕНСТВА, ПРИВЕДЕННОГО В П. 10.7.3

Пусть $f(\tau)$ и $g(\tau)$ — произвольные функции, в общем случае комплексные, от вещественной переменной τ, и пусть λ — вещественный параметр. Тогда

$$\int_{-\infty}^{+\infty} |f+\lambda g^*|^2 d\tau = \int_{-\infty}^{+\infty} (f+\lambda g^*)(f^*+\lambda g)\, d\tau \geqslant 0, \quad (1)$$

или

$$\int_{-\infty}^{+\infty} ff^*\, d\tau + \lambda \int_{-\infty}^{+\infty} (fg+f^*g^*)\, d\tau + \lambda^2 \int_{-\infty}^{+\infty} gg^*\, d\tau \geqslant 0. \quad (2)$$

Минимум этого выражения, квадратичного по λ, определяется дифференцированием, т. е.

$$\int_{-\infty}^{+\infty} (fg+f^*g^*)\, d\tau + 2\lambda \int_{-\infty}^{+\infty} gg^*\, d\tau = 0. \quad (3)$$

Корень $\lambda=\lambda_{\text{мин}}$ уравнения (3) равен

$$\lambda_{\text{мин}}=-\frac{1}{2}\frac{\int\limits_{-\infty}^{\infty}(fg+f^*g^*)\,d\tau}{\int\limits_{-\infty}^{+\infty}gg^*\,d\tau}. \tag{4}$$

Если подставить его в (2), то получим

$$4\left(\int\limits_{-\infty}^{+\infty}ff^*\,d\tau\right)\left(\int\limits_{-\infty}^{+\infty}gg^*\,d\tau\right)\geqslant\left(\int\limits_{-\infty}^{+\infty}(fg+f^*g^*)\,d\tau\right)^2. \tag{5}$$

Пусть

$$f=\tau\psi(\tau),\qquad g=\frac{d\psi^*(\tau)}{d\tau}. \tag{6}$$

Тогда

$$fg+f^*g^*=\tau\left(\psi\frac{d\psi^*}{d\tau}+\psi^*\frac{d\psi}{d\tau}\right)=\tau\frac{d}{d\tau}(\psi\psi^*), \tag{7}$$

и (5) после интегрирования правой части по частям и в предположении, что *) $\tau\psi\psi^*\to 0$ при $\tau\to\pm\infty$, принимает вид

$$4\left(\int\limits_{-\infty}^{+\infty}\tau^2\psi\psi^*\,d\tau\right)\left(\int\limits_{-\infty}^{+\infty}\frac{d\psi}{d\tau}\frac{d\psi^*}{d\tau}\,d\tau\right)\geqslant\left(\int\limits_{-\infty}^{+\infty}\psi\psi^*\,d\tau\right)^2. \tag{8}$$

Это и есть требуемое неравенство.

Равенство в (8) имеет место только в случае, когда выполняется равенство в (1), что возможно только при $f=-\lambda g^*$, или, используя (6), при

$$\frac{d\psi}{d\tau}=-\frac{1}{\lambda}\tau\psi. \tag{9}$$

Общее решение этого дифференциального уравнения имеет вид

$$\psi(\tau)=A\exp\left(-\frac{\tau^2}{2\lambda}\right), \tag{10}$$

где A — постоянная. Пригодны только решения с $\lambda\geqslant 0$, поскольку в противном случае $\psi(\tau)$ не будет обращаться в нуль на бесконечности. Следовательно, (8) *становится равенством тогда и только тогда, когда ψ есть функция Гаусса.*

ПРИЛОЖЕНИЕ 9

ВЫЧИСЛЕНИЕ ДВУХ ИНТЕГРАЛОВ, ИСПОЛЬЗУЕМЫХ В П. 12.2.2 **)

В настоящем приложении мы вычислим интегралы (12.2.8) и (12.2.9), т. е.

$$\mathscr{Y}_1=\frac{1}{4\pi}\iiint\limits_{V_1}\left\{\exp\left[i(px_1+qy_1)\right]\frac{\partial^2}{\partial z_1^2}\left(\frac{\exp\left(\frac{i\omega R}{c}\right)}{R}\right)\right\}dx_1\,dy_1\,dz_1, \tag{1}$$

$$\mathscr{Y}_2=\frac{1}{4\pi}\frac{\omega^2}{c^2}\iiint\limits_{V_1}\left\{\exp\left[i(px_1+qy_1)\right]\left(\frac{\exp\frac{i\omega R}{c}}{R}\right)\right\}dx_1\,dy_1\,dz_1, \tag{2}$$

*) Фактически это условие всегда выполняется, когда интегралы в левой части (8) сходятся (см. [46]).

**) По Дарвину [47].

где

$$R = +\sqrt{x_1^2+y_1^2+z_1^2} \tag{3}$$

и $\omega^2 > c^2p^2$. Рассмотрим два случая, а именно:

(а) $0 < y < d$, где y и d — постоянные, а область интегрирования V_1 представляет собой слой

$$-\infty < x_1 < \infty,$$
$$-y \leqslant y_1 \leqslant d-y,$$
$$-\infty < z_1 < \infty,$$

из которого исключена небольшая сфера исчезающе малого радиуса с центром в начале координат $x_1 = y_1 = z_1 = 0$.

(б) $y > d$ или $y < 0$; область интегрирования V_1 совпадает со всем слоем

$$-\infty < x_1 < \infty,$$
$$-y \leqslant y_1 \leqslant d-y,$$
$$-\infty < z_1 < \infty.$$

Для вычисления $\mathscr{Y}_1$ применим теорему Гаусса, записанную в виде

$$\iiint\limits_{V_1} \operatorname{div} \mathbf{G}\, dV_1 = \iint\limits_{S_1} \mathbf{G}\cdot\mathbf{n}\, dS_1, \tag{4}$$

где $\mathbf{G}$ — произвольная векторная функция координат, $\mathbf{n}(n_x, n_y, n_z)$ — единичный вектор внешней нормали к поверхности S_1, ограничивающей объем V_1. Положим

$$G_{x_1} = G_{y_1} = 0, \qquad G_{z_1} = \frac{1}{4\pi}\frac{\partial}{\partial z_1}\left[\frac{\exp\left[i\left(px_1+qy_1+\frac{\omega R}{c}\right]\right.}{R}\right], \tag{5}$$

тогда из (1) и (4) получим

$$\mathscr{Y}_1 = \frac{1}{4\pi}\iint\limits_{S_1} n_{z_1}\frac{\partial}{\partial z_1}\left(\frac{\exp\left[i\left(px_1+qy_1+\frac{\omega R}{c}\right]\right.}{R}\right) dS_1. \tag{6}$$

Поскольку $n_{z_1} = 0$ на обеих сторонах ($y_1 = -y$ и $y_1 = d-y$) слоя, то в случае (б) интеграл $\mathscr{Y}_1 = 0$ *). В случае (а) мы должны также учесть вклад от небольшой сферы σ_1 с центром в начале координат. Если a — радиус сферы, то $n_z = -z_1/a$, и этот вклад равен

$$\frac{1}{4\pi}\iiint\limits_{\sigma_1}\left(\frac{z_1}{a}\right)^2 \exp\left[i\left(px_1+qy_1+\frac{\omega a}{c}\right)\right]\left(\frac{1}{a^2}-\frac{i\omega}{ac}\right)d\sigma_1 \to \frac{1}{3} \quad \text{при } a\to 0. \tag{7}$$

Следовательно,

$$\left.\begin{aligned} \mathscr{Y}_1 &= \frac{1}{3}, \quad \text{когда } 0 < y < d, \\ \mathscr{Y}_1 &= 0, \quad \text{когда } y < 0 \text{ или } y > d. \end{aligned}\right\} \tag{8}$$

Прежде чем вычислять $\mathscr{Y}_2$, отметим, что интегрирование по исчезающе малой сфере с центром в начале координат не дает никакого вклада, поскольку подынтегральное выражение содержит только особенность порядка $1/R$. Поэтому в случае (а), как и в случае (б), нужно интегрировать

*) Строго говоря, поверхность S_1 должна быть замкнутой; поэтому необходимо рассматривать также вклады от удаленных краев слоя. Однако этими вкладами можно пренебречь по физическим соображениям, так как их эффекты проявятся в исследуемых точках через бесконечно большой промежуток времени.

по всему объему слоя, т. е.

$$\mathscr{Y}_2 = \frac{\omega^2}{4\pi c^2} \int\limits_{-y}^{d-y} \exp(iqy_1)\, \mathscr{L}(y_1)\, dy_1, \tag{9}$$

где

$$\mathscr{L}(y_1) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} \frac{\exp\left[i\left(px_1 + \frac{\omega R}{c}\right)\right]}{R}\, dx_1\, dz_1. \tag{10}$$

Выберем в (10) новые независимые переменные ρ и χ, определяемые соотношениями

$$px_1 + \frac{\omega}{c} R = \rho \sqrt{\left(\frac{\omega}{c}\right)^2 - p^2}, \tag{11а}$$

$$z_1 = \sqrt{\rho^2 - y_1^2}\, \sin\chi, \tag{11б}$$

где корни в (11а) и (11б) берутся со знаком плюс. На плоскости $x_1 z_1$ кривые $\rho = \text{const}$ имеют вид эллипсов, а χ — угол между радиусом-вектором и большой осью эллипса. При $\rho = |y_1|$ эллипс вырождается в точку. Таким образом, всей плоскости $x_1 z_1$ соответствуют следующие интервалы значений ρ и χ: $|y_1| \leqslant \rho \leqslant \infty$, $0 \leqslant \chi \leqslant 2\pi$. Следовательно,

$$\mathscr{L}(y_1) = \int\limits_0^{2\pi} \int\limits_{|y_1|}^{\infty} \frac{\exp\left[i\rho \sqrt{\frac{\omega^2}{c^2} - p^2}\right]}{R} \frac{\partial(x_1, z_1)}{\partial(\rho, \chi)}\, d\rho\, d\chi, \tag{12}$$

где

$$\frac{\partial(x_1, z_1)}{\partial(\rho, \chi)} = \frac{\partial x_1}{\partial\rho} \frac{\partial z_1}{\partial\chi} - \frac{\partial x_1}{\partial\chi} \frac{\partial z_1}{\partial\rho} \tag{13}$$

является якобианом нашего преобразования. Из (3) и (11а) находим

$$\left.\begin{aligned} \frac{\partial x_1}{\partial\rho} &= \frac{\sqrt{\left(\frac{\omega}{c}\right)^2 - p^2}}{p + \frac{\omega}{c}\frac{x_1}{R}}, \\ \frac{\partial x_1}{\partial\chi} &= 0, \end{aligned}\right\} \tag{14а}$$

а из (11б) имеем

$$\left.\begin{aligned} \frac{\partial z_1}{\partial\rho} &= \frac{\rho}{\sqrt{\rho^2 - y_1^2}} \sin\chi, \\ \frac{\partial z_1}{\partial\chi} &= \sqrt{\rho^2 - y_1^2}\cos\chi = \sqrt{\rho^2 - y_1^2 - z_1^2}. \end{aligned}\right\} \tag{14б}$$

Кроме того, справедливо тождество

$$\left(pR + \frac{\omega}{c} x_1\right)^2 - \left(px_1 + \frac{\omega}{c} R\right)^2 = \left[\left(\frac{\omega}{c}\right)^2 - p^2\right] [x_1^2 - R^2],$$

или, используя (3) и (11а),

$$\left(pR + \frac{\omega}{c} x_1\right) = \sqrt{\left(\frac{\omega}{c}\right)^2 - p^2}\, \sqrt{\rho^2 - y_1^2 - z_1^2}. \tag{15}$$

Из (13), (14) и (15) следует, что

$$\frac{\partial(x_1, z_1)}{\partial(\rho, \chi)} = R. \tag{16}$$

Подставим теперь (16) в (12). Интегрирование по χ дает сразу же множитель 2π. Интегрирование по ρ тоже не представляет труда, и мы

получим

$$\mathcal{L}(y)=\frac{2\pi i \exp\left[i\,|\,y_1\,|\,\sqrt{(\omega/c)^2-p^2}\right]}{\sqrt{(\omega/c)^2-p^2}}, \tag{17}$$

где на основании таких же физических соображений, как и ранее, мы отбросили осциллирующий вклад от бесконечности.

Подставляем теперь (17) в (9) и вычисляем получившийся интеграл. Имеем

$$\left.\begin{aligned}
\mathcal{Y}_2&=\left(\frac{\omega}{c}\right)^2\frac{\exp(-igy)}{2g\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}}\{\exp(igd)-1\}, \quad \text{когда } y>d,\\
\mathcal{Y}_2&=\left(\frac{\omega}{c}\right)^2\frac{\exp(-ihy)}{2h\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}}\{\exp(ihd)-1\}, \quad \text{когда } y<0,\\
\mathcal{Y}_2&=\left(\frac{\omega}{c}\right)^2\left\{\frac{1}{gh}-\frac{\exp(-igy)}{2g\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}}+\frac{\exp[-ih(y-d)]}{2h\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}}\right\},\\
&\qquad\qquad \text{когда } 0<y<d,
\end{aligned}\right\} \tag{18}$$

где

$$g=q-\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}, \quad h=q+\sqrt{\left(\frac{\omega}{c}\right)^2-p^2}. \tag{19}$$

ЛИТЕРАТУРА

1. R. Courant, Differential and Integral Calculus, Blackie a. Son, Glasgow, 1936. (Р. Курант, Курс дифференциального и интегрального исчисления, ГТТИ, 1931.)
2. K. Schwarzchild, Königl. Gess. Wiss. Göttingen, Math.-Phys. Kl., 3, 126 (1903).
3. W. Ehrenberg, R. E. Siday, Proc. Phys. Soc. B62, 8 (1949).
4. H. Boersch, Naturwissenschaften 28, 709 (1940).
5. W. Glaser, Grundlagen der Elektronenoptik, Springer, Wien, 1952, p. 548.
6. P. Debye, Math. Ann. 67, 535 (1909).
7. H. Poincaré, Acta Math. 8, 295 (1886).
8. E. Jeffreys, B. S. Jeffreys, Methods of Mathematical Physics, Cambr. Univ. Press, 1946, ch. 17.
9. A. Erdélyi, Asymptotic Expansions, Dover Publications, New York, 1956.
10. G. N. Watson, Theory of Bessel Function, Cambr. Univ. Press, 2nd ed., 1948, ch. 8. (Г. Ватсон, Теория бесселевых функций, ИЛ, 1949, гл. 8.)
11. C. Chester, B. Friedman, F. Ursell, Proc. Cambr. Phil. Soc. 53, 599 (1957).
12. W. Pauli, Phys. Rev. 54, 925 (1938).
13. H. Ott, Ann. d. Physik 43,, 393 (1943).
14. P. C. Clemmow, Quart. J. Mech. Appl. Math. 3, 241 (1950).
15. P. C. Clemmow, Proc. Roy. Soc. A205, 286 (1951).
16. B. L. van der Waerden, Appl. Sci. Rev. B2, 33 (1952).
17. W. Thomson, Proc. Roy. Soc. A42, 80 (1887).
18. G. N. Watson, Proc. Cambr. Phil. Soc. 19, 49 (1918).
19. T. J. I'A. Bromwich, An Introduction to the Theory of Infinite Series, Macmillan, London, 1908, p. 447.
20. J. Focke, Ber. Sachs. Ges. Akad. Wiss. 101, H. 3 (1954).
21. J. G. van der Corput, Indag, Math. 10, 201 (1948).
22. D. S. Jones, M. Kline, J. Math. Phys. 37, 1 (1958).
23. N. Chako, C. R. Acad. Sci. 247, 436, 580, 637 (1958).
23a. M. Kline, I. W. Kay, Electromagnetic Theory of Geometrical Optics, New York, Intersci. Publ., N. Y., 1965, ch. XII.
24. N. G. van Kampen, Physica 14, 575 (1949).
25. G. Salmon, Analytic Geometry of three Dimensions, Vol. I, rev. by R. A. P. Rogers, Logmans, Green a. Co., London, 5th ed., 1912, p. 411.
26. P. Dirac, The Principles of Quantum Mechanics, Clarendon Press, Oxford, 1930. (П. А. М. Дирак, Принципы квантовой механики, Физматгиз, 1960.)
27. B. van der Pol, H. Bremmer, Operational Calculas based on the two-sided Laplace Integral, Cambr. Univ. Press, 1950, pp. 62, 100.

28. L. Schwartz, Théorie des distributions, Hermann et Cie, Paris, 1950, Vol. 1, 1951, vol. II.
29. G. Temple, Proc. Roy. Soc A228, 175 (1955).
30. M. J. Lighthill, An Introduction to Fourier Analysis and Generalised Functions, Cambr. Univ. Press, 1958.
31. A. E. H. Love. Proc. Lond. Math. Soc. 1, 56 (1904).
32. H. Bateman, Partial Differential Equations of Mathematical Physics, Cambr. Univ. Press, 1932, p. 196.
33. T. Levi — Civita, Caractéristiques des Systémes Différentiels et Propagation des Ondes, Librairie Felix Alcan, Paris, 1932, § 10.
34. R K. Luneburg, Mathematical Theory of Optics, Brown Univ., 1944, p. 21; печатное издание Univ. California Press, Berkeley a. Los Angeles, 1964, § 6, 7.
35. M. Kline, Comm Pure a. Appl. Math. 4, 239 (1951).
36. A. Rubinowicz, Acta Phys. Polonica 14, 209 (1955).
37. H. Bremmer, Comm. Pure a. Appl. Math. 4, 419 (1951).
38. E. T. Copson, Comm. Pure a Appl. Math. 3, 427 (1951).
39. F. Zernike, Physica 1, 689 (1934).
40. F. Zernike, H. C. Brinkman. Verh. Akad. Wet. Amst. (Proc. Sec. Sci.) 38, 161 (1935).
41. B. R. A. Nijboer, Thesis, Univ. of Groningen, 1942.
42 A. B. Bhatia, E. Wolf, Proc. Cambr. Phil. Soc. 50, 40 (1954).
43. M. Born, Natural Philosophy of Cause and Chance, Clarendon Press, Oxford, 1949, p. 153; Dover Publications, New York, 1964, p. 153.
44. B. Courant, D. Hilbert, Methods of Mathematical Physics, Intersci. Publ. New York, 1953, Vol. 1. (Р. Курант, Д. Гильберт, Методы математической физики, Гостехиздат, 1951, т. I.).
45. E Kemble, The Fundamental Principles of Quantum Mechanics, McGraw-Hill, New York, 1937, p. 594.
46. H. Weyl, The Theory of Groups and Quantum Mechanics, Methuen, London, 1931.
47. C. G. Darwin, Trans. Cambr. Phil. Soc. 23,, § 6, 8 (1924).

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Макс Борн, Эмиль Вольф

ОСНОВЫ ОПТИКИ

М., 1973 г., 720 стр. с илл.

Редактор *Н. А. Райская*

Техн. редактор *И. Ш. Аксельрод* Корректор *Л. Н. Боровина*

Сдано в набор 24/IV 1973 г. Подписано к печати 20/VIII 1973 г. Бумага $70\times108^1/_{16}$. Физ. печ. л. 45+вкл. 0,25 п. л. Условн. печ. л. 63,35. Уч.-изд. л. 66,33 Тираж 20 000 экз. Цена книги 5 р. 10 к. Заказ № 301

Издательство «Наука»

Главная редакция физико-математической литературы

117071, Москва В-71, Ленинский проспект, 15

Ордена Трудового Красного Знамени

Первая Образцовая типография имени А. А. Жданова

Союзполиграфпрома при Государственном комитете

Совета Министров СССР по делам издательств,

полиграфии и книжной торговли

Москва М-54 Валовая, 28